

Ducking big problems is cowardly

Last week's meeting of seven industrialized governments in Munich accomplished little, but rather is a saddening reminder that even the most powerful governments are sometimes mesmerized by the voters' power.

POLITICIANS are skilled at the art of compromise, right? And the more powerful politicians are, the more able they are to wring compromises from circumstances too complex for lesser mortals to resolve, right? So what then is to be made of the failure to decide anything of substance by the seven most powerful politicians of the industrialized countries at their meeting in Munich last week? True, 'G-7', as it likes to be known, agreed that Mr Boris Yeltsin might attend a dinner to which he had not been invited, and went on to confirm its earlier decision that Russia and the other ex-Soviet republics could enjoy financial credits amounting to \$24 billion once the International Monetary Fund is satisfied with the progress of reform in Russia (which may happen by the autumn); meanwhile, there is to be a payment of \$1 billion on account. Otherwise, important issues were mostly dodged.

The chief casualty of this indecision is the attempt to negotiate a new version of the General Agreement on Tariffs and Trade (GATT), on which governments have been labouring since 1987, and which should have been completed last December. A year ago at Houston, G-7 declared that it would do everything in its power to see the negotiations completed by the deadline. Now, it is merely "confident" that the negotiations can be completed by the end of this year. Why has the ardour so cooled?

The immediate difficulty seems to have been that President François Mitterrand, whose farmers have already been kicking up a fuss over the reform of Europe's Common Agricultural Policy, itself a necessary ingredient of a GATT deal, is reluctant to give the issue further publicity in advance of the French referendum on the Maastricht Treaty arranged for September. But if Mitterrand had not pleaded for caution, President George Bush would no doubt have done so. For is he not about to embark on an election campaign in which defence of US jobs against the depredations of overseas trading partners will be necessary for success?

The irony is that G-7's chief concern is, or should be, the economic well-being of the world, that its members are also in different ways all worried about their own economies — and that a successful GATT agreement is the only obvious way of putting them to rights. For more than 30 years, the fastest growing component of the world's economy has been the growth of international trade, whose volume has almost steadily increased by about 10 per cent a year, faster even than domestic economic growth in, say, Japan.

This is almost entirely a consequence of understandings such as GATT, which among other things regulates allowed

levels of import tariffs and which prevents a member government from discriminating among its trading partners. (Yeltsin's wish that Russia should join GATT is simply a wish that it should no longer be discriminated against.) Now, people are estimating that a successful GATT agreement would add 2 per cent to the growth rate of the G-7 members as well as taking the edge off the economic plight of some of the world's poorest countries. It is monumentally shortsighted, perhaps even electorally, to let that benefit be postponed simply because there are some ticklish public votes early on the agenda of some G-7 members.

Reflection on its own agenda at Munich should otherwise have persuaded G-7 that there is no other way out of the economic box in which its members find themselves. The credits being negotiated with Russia and the other members of the Commonwealth of Independent States (CIS) are huge, but probably insufficient to see these countries through to stability. (Simply winning German reunification will cost Germany 5 per cent of its gross domestic product for at least a decade.)

To the extent that much of the Russian money will be spent on imports from the countries that provide the credits, the result will help to stimulate Western economies, but only at the risk of inflation, just as if the governments concerned paid people for digging holes in the ground and then for filling them in again. The productive use of the CIS credits, likely to entail joint investment projects with companies in the West, will entail real investments and thus a drain on what companies can afford to invest domestically. That is why it is inevitable that the politically, socially and culturally admirable goal of helping the countries of the CIS to escape their past must be a brake on economic growth elsewhere. That is why the recession of the past two years is so obdurate. In the circumstances, to let the benefits of GATT be postponed even by six months is foolhardy, to say the least of it. □

Al Gore's future

A science-minded politician has appeared on the Democrats' ticket for the US election — in the nick of time.

PAUL Tsongas, the first presidential candidate in this year's US general election to capture attention for the technical content of what he had to say (see *Nature* 351, 427; 1991), has long since bitten the dust. Although he captured pri-



mary votes in a handful of states on a scale that appeared to surprise him, he eventually ran out of funds. But the selection last week of Senator Albert Gore, the Tennessee Democrat, as the Democratic party's candidate for vice-president in November's election will put another science-minded politician on every ballot.

Gore, generally known as "Al" and sometimes as "Senator Science", has an impressive record in the Congress, first in the House of Representatives and, since 1984, in the Senate. In the House he was influential, more than a decade ago, in drawing attention to the corrosiveness of fraud in science. Now, witnesses before the Senate's science subcommittee he chairs are repeatedly challenged by his knowledge of issues such as global warming, genetic engineering and high-speed computer networks. So does that mean that the election campaign, traditionally supposed to start at the end of August, but effectively under way from the Democrats' convention in New York this week, will break new ground by giving prominence to technical issues?

Sadly, no. The traditional function of a candidate for vice-president is to do his running-mate's dirty work, attacking the intelligence, honesty and character of the opposing candidate for president, allowing his partner to seem to stand above the slanging-match. Gore seems well-fitted for the role: he has a sharp wit, a good constitution and experience of how mud is slung from his own unsuccessful quest for the Democratic nomination in 1988. Gore may not have time to say much about science until after November.

But then, he could be influential. The Democrats' candidate for President, William ("Bill") Clinton, the governor of Arkansas, has said that, if he is elected president, he will turn Gore loose on such issues as the environment, arms control and technological innovation. On all these issues, Gore is best described as a hard-headed liberal. More controversially, he also shares the Democrats' current belief that the United States needs what is called an 'industrial policy', understood to mean government support for advanced technical innovation. Gore has been a strong exponent of this view. It would be instructive for the rest of us if Clinton and Gore were given a chance to show that the United States can succeed where Britain and France have failed. The United States has a splendid record in carrying high-technology into weapons systems, chiefly through the Defense Advanced Research Projects Agency (DARPA), but civilian life is faster-moving. The lesson from Japan, which Britain now seeks to learn, is that governments function best as partly fond uncles, partly goads, of the industries they sponsor.

Two important 'ifs' hang over this prospect. First, Clinton would have to be elected. Second, he would have to keep his promise that his vice-president will have more useful things to do than to hang around for the occasions when the Senate is equally divided in a vote so as to exercise the administration's right to break the tie. Incumbents at the White House have a long record of promising that *their* vice-presidents will not be just pretty faces (which has not always been the case), and then of disappointing them. But it could be fun if Clinton wins and keeps his promise. □

Patents for genomes?

The United States could (and should) do more to prevent the confusion caused by genome patenting.

THE great brouhaha on the patenting of nucleotide sequences fished from the human genome is taking on elements of farce. When the US National Institutes of Health (NIH) decided last year to seek patent protection for sequences published by Dr Craig Venter and his associates, it failed to guess what the consequences would be. In Britain, one has been the Medical Research Council's decision to protect its interests by following suit, and meanwhile to keep its own genome database to itself. (Under British patent law, there is no grace period of 12 months between publication and the application for a patent such as applies in the United States.) Meanwhile, the governments of Britain, France and Japan have deplored NIH's policy, and have sought an international understanding that sequences whose function is unknown should not be patentable. Of the damage done to the civility of the international genome enterprise, perhaps the most serious is the loss of Dr James Watson from the US project, chiefly because of his siding with the angels on this question.

But now, it seems, the United States is playing games. Dr D. Allan Bromley, the president's science adviser, is reported (see page 180) to believe that the NIH patent applications will not succeed because most of the sequences (usually those of parts of genes) are not accompanied by an understanding of their function. That, he says, is why an international agreement is neither necessary nor desirable. When the patent applications are rejected by the US Patent Office, the argument runs, an international agreement would be redundant. And Bromley believes the Patent Office will on this occasion move more quickly than is its habit.

This is an odd way to deal with such an important issue, whose unwelcome influence on the civility of the international research enterprise has already been significant. NIH takes the line that its public duty requires it to seek to patent everything patentable, if only for the (good) reason that it can then ensure that worthwhile developments are vigorously exploited. But does not NIH, a government agency, owe its public duty to the US government? If the government asked NIH to withdraw the patent applications, it would have to do so. And if the president's science adviser believes the attempt to patent DNA sequences will end in failure, is not he or the administration as a whole able to stop what seems a wild-goose chase before it does more damage?

As it happens, the issue could usefully have been brought up at last week's summit meeting of G-7 (see above). A few years ago, the heads of government used these occasions to initial agreements on international science projects in need of high-level impetus. All the important participants in the genome row were at Munich. Why did none of them mention it? Then, at least, there might have been something to boast of. □

US drops Imanishi-Kari investigation; Baltimore withdraws *Cell* retraction

Washington. US prosecutors announced this week that they are dropping their criminal investigation of Thereza Imanishi-Kari, a Tufts University immunologist accused of scientific misconduct, because the case is too complex and tangled for a jury to reach a conclusion. At the same time, David Baltimore, a Nobel prizewinner and co-author with Imanishi-Kari of the questioned work, which appeared in *Cell* in 1986, said that he would withdraw a retraction made last year after the second of two investigations by the National Institutes of Health (NIH) found evidence of misconduct. Although the government action is a victory for Imanishi-Kari, the saga, which has lasted six years and become the centrepiece for debate over research ethics, is expected to continue.

Baltimore resigned last year from the presidency of Rockefeller University after being dogged by criticism of his conduct during the extended investigation. He says that his retraction was made on the basis of statistical and forensic studies commissioned by NIH that appeared to prove that Imanishi-Kari had fabricated data.

But an independent analysis commissioned earlier this year by Imanishi-Kari's attorney, Bruce Singal of the Boston law firm Ferriter, Scobbo, Sikora, Caruso & Rodophele, concluded that the forensic analysis by the US Secret Service was seriously flawed. The reanalysis, by Albert Lyter, a former US Treasury forensic chemist, concluded that the Secret Service analysis suffered from "inadequate samples of data, erroneous interpretations of test results . . . and poor methodology and procedures."

Singal says that all he and Imanishi-Kari needed was a chance to see the data. "We've been crying in the wilderness for years, trying to get the data [on which the Secret Service based its conclusions]", he says. "Finally we got a responsible party in the form of the US attorney, who allowed us to see the data."

Baltimore says there "has been no contradiction and extensive confirmation" of the *Cell* paper, including replication of the central experiment by Imanishi-Kari herself. "I now believe that there is no reason for the paper to be in limbo", he says. The new forensic analysis "is a confirmation of my feeling all along that Imanishi-Kari had not attempted to deceive anybody".

However, Geoffrey Garinther, an assistant US attorney in Baltimore, Maryland, says that the reanalysis of the forensic evidence did not cause his office to end its 18-month investigation. "Our decision was not based on a lack of confidence in the Secret Service findings," he says. Instead, he says

that prosecutors felt that the controversy was essentially a scientific one and that a conviction would be extremely difficult.

"We couldn't take this to the jury and ask them to decide it when there are so many scientists who disagree", he says. If Imanishi-Kari's lawyers "could suggest to the jury that she'd done the experiments or that her results had been verified by others, that's enough for reasonable doubt", he says.

Phillip Sharp, a Massachusetts Institute of Technology researcher who has championed Baltimore's case, says that Baltimore had never wanted to retract the paper, and jumped at the first opportunity to reverse his decision. Sharp says that the retraction in March 1991 had been "under the pressure of being president of Rockefeller and wanting to put the whole thing behind him".

In a letter dated 13 July to the oversight and investigations subcommittee of the Energy and Commerce Committee, chaired by US Representative John Dingell (Democrat, Michigan), Richard Bennett, the US attorney for Maryland, pointed out that the statistical evidence upon which NIH had relied would be inadmissible in a criminal trial. And he predicted that the Secret Service's forensic analysis would have been contradicted by the independent analysis commissioned by Singal. "We are confident that the [Secret Service] testimony would have been persuasive," he wrote. "Our duty, however, is not simply to persuade, but to persuade beyond a reasonable doubt."

A grand jury had already heard from

witnesses when prosecutors decided not to seek an indictment in the case. The jurors "were clearly incapable of understanding" the case, says one scientist who testified before the grand jury and who did not wish to be identified. "They literally pulled 24 people off the street."

Dingell said that "the decision not to prosecute does not change the fact that the *Cell* paper was retracted because of serious, and extensive, irregularities". He said that the US attorneys had told his subcommittee that they agreed with the Secret Service finding of falsification of data "but believe that the matter was too complicated for a lay jury".

Suzanne Hadley, the former NIH investigator who wrote the report that found evidence of misconduct, criticized the decision to drop criminal prosecution. "If [Imanishi-Kari] did falsify data, or lie to federal investigators, those were federal crimes and they should be dealt with in the criminal sector", she says. The NIH investigation is continuing, although Hadley was removed from her position after clashing with NIH officials over her impartiality.

Imanishi-Kari described the six-year inquiry as a "terrible injustice". Now that she no longer faces the threat of criminal prosecution, she says she will ask NIH to release a 1989 grant that had been frozen during the investigation. In the wake of the decision, she says, "I feel that, after all that, there is some justice. I was doubting before."

Christopher Anderson & Traci Watson

Audit faults US national labs

Washington. As if the misuse of federal research funds by prominent universities was not bad enough, the US government now believes that some of its own laboratories have been playing fast and loose with federal money. Last week, the US Senate released a report accusing some of the government's 39 research centres of inadequate accounting, careless auditing and poor cost control. Such practices, it says, have "contributed to the wasteful or inappropriate use of millions of federal dollars".

The report by an oversight committee of the Senate Committee on Government Affairs says that poor record-keeping prevents it from knowing how much money has been misspent. The Lawrence Livermore, Los Alamos and Lawrence Berkeley national laboratories — all operated by the University of California for the Department of Energy (DOE) — are the most frequently

mentioned in the report.

The report deals with some of the same management issues that are being investigated by Congress as part of an inquiry into how universities bill the government for the overhead cost of conducting research. That investigation has uncovered millions of dollars in improper charges.

Among the most egregious examples of the University of California's behaviour are the following: From 1988 to 1991, the DOE reimbursed the university for the cost of leasing 58 vehicles for the Livermore site, although the department had ordered Livermore to reduce the size of its fleet. And the university continued to collect lease money even after the DOE bought the vehicles and gave them to the university. Tommy Ambrose, who manages UC's office of laboratory affairs, agrees that "we didn't manage that situation well".

Traci Watson

Researchers see obstacles in using spy data

Washington. For decades, the US military has used aircraft, spy satellites and nuclear submarines to gather classified data about features on land and in the oceans. Now, prodded by politicians, the military has begun to release some of its long-guarded information in the hope that it will benefit a range of scientists, in particular those studying global climate change. Ironically, this decision comes just as researchers are beginning to obtain similar information from nonmilitary sources in other countries.

Even as scientists welcome the new openness, many researchers question the value of the data. Although such material as satellite measurements of the ocean's surface is expected to help scientists to learn more about the topography of the ocean floor and seafloor plate tectonics, much of the data seem likely to be inappropriate or incomplete.

The military's efforts to distribute once-secret information began in 1990, when Senator Al Gore (Democrat, Tennessee) learned that scientists were looking for evidence of global warming by studying changes in polar ice thickness. The available ice data were sketchy, so Gore asked the US Navy to release ice data collected during Arctic submarine voyages. The Navy complied, declassifying sonar measurements of ice thickness taken between 1977 and 1989. Since then, the Navy has also declassified satellite data about the ocean surface near the South Pole and approved the release of other ocean surface data.

This year, at the urging of Gore and Senator Frank Murkowski (Democrat, Alaska), the Central Intelligence Agency (CIA) agreed to open its doors to a panel of scientists, not yet chosen, seeking data that could benefit environmental research. And in June, the White House called on the CIA and the Department of Defense (DoD) to make public any "appropriate" data about global climate change.

Scientists say they are always eager to scrutinize new data, no matter what the source. And a few researchers have longed for access to particular military datasets, which they feel certain will be of use. It is thought, for example, that the military holds records dating back several decades of ocean temperatures, forest acreage and cloud cover. If these data are continuous and easily analysed, they may benefit studies of global climate change.

But others are less convinced that the declassified information will contribute to their work. Even the catalogues describing reconnaissance data have been kept secret. Nor is anything known about the quality of

the data. The information was not collected for scientific use, and it may lack the necessary consistency and reliability.

Scientists are especially concerned about the usefulness of satellite data. The military needs satellite pictures with high resolution and high contrast — to look for tanks, for example. But most scientists want accurate absolute values, which can be difficult to record with military systems.



The nuclear submarine *Seahorse* at the North Pole.

"To me, it may be the temperature of the Gulf Stream that's important," says Otis Brown, associate dean of research at the Rosenstiel School of Marine and Atmospheric Science in Miami, Florida. "To the military, it may be the front of the Gulf Stream — the edge of the feature — that's important."

Even for images taken at an appropriate resolution, the form in which images are released could negate their scientific value. For instance, satellites often record cloud cover in terms of voltage differences. But data in that form are so abstract as to be useless unless they are first processed.

Unfortunately, data processing takes time. While it is usually handled by federal agencies, in particular the National Aeronautics and Space Administration (NASA) or the National Oceanic and Atmospheric Administration (NOAA), it is not clear that either would be willing to commit the resources needed to convert military data into a form that scientists can use.

The reconnaissance data that have

already been declassified send mixed signals about the usefulness of data yet to come. On the one hand, information from the Navy's Geodesy satellite, launched in 1985 to measure the height of the ocean surface, has proved a boon to many scientists. The data, processed before their release to researchers, were accompanied by information about the satellite. Geodesy's ocean surface data are now helping seafloor topographers to draw a profile of the previously unknown Southern Ocean basin. Geodesy's information is also a windfall for plate tectonics researchers, who can use the sea-surface data to study the underwater edges of the Earth's crustal plates.

On the other hand, the ice-thickness data collected aboard Navy nuclear submarines have so far proved problematic. The submarine's movement can introduce unpredictable error into ice profiles, as can the instruments that measure the ship's depth and the underwater thickness of the ice pack.

Carl Wales, special projects officer at the Submarine Laboratory in San Diego, California, says that a submarine's ice-measuring equipment "works well as an operating instrument. But no credible scientist would allow you to sell it to him as a scientific device. It wasn't designed to be one."

Climate scientists may eventually obtain their ice-thickness data from another source. Administrators at the Woods Hole (Massachusetts) Oceanographic Institution are weighing an offer made last year by Russian officials to convert one of their nuclear submarines into a scientific vessel. A

Russian research submarine that crisscrossed the entire Arctic could collect more and better data than the US Navy now possesses.

The Navy's Geodesy data also seem likely to be superseded in the next decade. The European Remote Sensing 1 satellite, launched in 1991 by the European Space Agency, will collect essentially the same data that the Navy has released — as well as data it has not been willing to release. The TOPEX/Poseidon satellite, a joint mission between NASA and France's Centre National d'Etudes Spatiales, is scheduled to be launched in August and will also yield much of the data the Navy has declassified. By releasing the Geodesy data now, the Navy is giving US researchers a head start on their European counterparts.

It is too soon to know what the US military will eventually serve up to scientists. But it seems likely that much of the military's masses of data will end up on the shelf, useful only for their intended military purposes rather than for basic research.

Traci Watson

British military beckons environmental researchers

London. The British Ministry of Defence (MoD) is putting its people, equipment and real estate at the disposal of environmental researchers in a new initiative intended to benefit science. Researchers welcome the opportunity but are scratching their heads to understand such generosity.

When the MoD's chief scientist, Sir Ron Oxburgh, announced the initiative two weeks ago, he spoke of technological interchange and the need for the services to know about

before the armed forces arrived. More surprising to Sutton was the military's scientific contributions. The expedition benefited from Royal Air Force officers with expertise in ornithology, a major in the Royal Marines with skills in archaeology and a member of the Royal Navy with a background in biological computation.

The information gathered was shared with the Belize department of natural resources and with the British Overseas Development

Agency, and it will also be given to other neotropical research groups. A second expedition is planned.

This enthusiasm for science has already made possible, and at times even initiated, many projects. By providing a formal application procedure, the new joint effort should greatly increase the amount of research conducted.

The MoD owns more than 600,000 acres of land in Britain, including 200 or so areas that are designated sites of special scientific interest (SSSI). Some of these ecosystems are in pristine condition, and have not been ploughed, sprayed or exposed to fertilizer for many decades.

The presence of the military also means that researchers can be more confident about leaving expensive monitoring equipment on these sites for long periods of time.

Researchers have had their eye on several pieces of property owned by the military. For ecologists looking for an example of chalk grassland to be included in the Ecological Change Network, a long-term environmental monitoring programme, there are several good examples on Salisbury Plain, the MoD's training ground in the south of England. Similarly, the gunnery ranges at Cape Wrath in the far north of Scotland hold great promise for ornithologists.

More important is the prospect that the military will allow its resources and personnel to be used by researchers. This means access to remote or hostile environments, allowing monitoring equipment to hitch rides on ships and planes, or using military personnel to make routine measurements.

Although the MoD cannot subsidize any projects, researchers believe that it should be relatively easy to tailor projects to fit the forces' routine operations or scheduled training activities.

Ian Mundell

Giotta encounters second comet; ready for more

Munich. Reactivated following its damaging encounter with Halley's comet in 1986, the European space probe Giotto surprised and delighted scientists last Friday in a successful encounter with the dying Grigg-Skjellerup comet. Passing within an estimated 200 km of the comet's nucleus, Giotto has not only provided important and high-quality data but has also emerged unscathed and ready for a third mission before the end of the century.

Launched in 1985 and designed to end its life after meeting Halley, Giotto provided researchers with a bonus second mission at a tiny cost. Now there is the possibility of a third, as yet undefined, encounter, making the ECU145-million (US\$190 million) space probe particularly good value for money.

The probe is no longer operating at full strength. Two of its ten scientific instruments — the neutral mass spectrometer and the particle impact analyser — and parts of the probe's extensive protective shield were destroyed in the Halley fly-by. The optical path of the camera was also blocked when hit by particles, although not before the camera had recorded the first pictures ever seen of a comet nucleus. After Giotto left Halley, scientists at the mission control centre in Darmstadt, Germany, pointed the probe towards the Earth before 'switching it off' for four years.

In February 1990, Giotto was reactivated, at a cost of ECU12 million, and aimed at its second target, a comet named for its two co-discoverers, New Zealander John Grigg and Australian John Francis Skjellerup. The comet is often upset by the gravitational pull of Jupiter, and the previous such disturbance, in 1964, pulled it just inside the Earth's orbit and within reach of Giotto. Unlike the fresh and active Halley, Grigg-Skjellerup is eroded and barely active, providing an interesting contrast for studies of the common origin of comets.

Last week's experiments concentrated on plasma physics, and preliminary results contain some surprises. For example, the 'bow shock' — the shock front that precedes a comet — was much stronger than expected, and was measurable at a much greater distance, around 16,000 km, than predicted. Magnetic wave phenomena around the comet were also unexpected, although they were textbook clear, according to Gerhard Schwehm, project scientist of the Giotto Extended Mission.

A third mission is now being discussed for 1999, and manoeuvres to set Giotto on its path back towards Earth have been delayed by a few weeks to conserve fuel.

Allison Abbott



British soldiers unload supplies from a helicopter in support of a joint scientific exercise in Belize.

the environments in which they operate. But researchers who have worked with the military before suggest that the explanation may be as simple as satisfying the wishes of senior officers who like science.

The project, called Joining Forces for the Environment, is being undertaken with the Natural Environment Research Council. It provides a formal structure under which researchers can apply to work at sites on MoD land or ask the military for help in carrying out scientific projects. Although it applies to any of the sciences, the chief beneficiaries are expected to be environmental researchers and ecologists.

David Sutton, a botanist at the Natural History Museum, London, involved in co-ordinating the scientific element of a military training exercise to Belize in the Yucatan Peninsula of Central America, has been impressed with the commitment and expertise in the services. "There are a lot of skills that service people have that are useful to scientists — in logistics, communications and navigation", Sutton says.

The Belize expedition, which took place in January and February 1991, made a study of the botany, ecology, zoology and soil science of the upper reaches of the Rospaculo River. This area had never been visited for scientific purposes and was inaccessible

Wall Street remains bearish on value of genome project

Washington. Investors are unimpressed by last week's announcement (see *Nature* 358, 95; 1992) that a venture fund is starting a \$70-million company to commercialize genome project research. Unresolved patent issues, limited markets for gene-based diagnostics and the slow progress of technology still make the genome project a risky prospect in their minds, despite widespread recognition of the long-term potential. Although the picture could change dramatically with a favourable decision by the US Patent Office on cDNA patents or the discovery of a key gene, few investors are ready to gamble their money on its prospects.

The Institute for Genomic Science and Human Genome Science Inc. — the research and marketing arms of the new company — will initially focus on sequencing cDNA, copies of genes that are expressed in the human body. Analysts expect that to be one of the few marketable areas of genome research in the near term.

Although researchers do not know the function of most of the genes that match the cDNA sequences, they can at least tell investors that the area is worth pursuing. The same cannot be said about DNA sequencing in general, where the chance of identifying a disease gene or other useful DNA regions is orders of magnitude lower and the slow pace of sequencing technology development is most keenly felt.

It may be that there is room for only one big company. J. Craig Venter, the former US National Institutes of Health (NIH) researcher around whom the new venture has been founded, may finish sequencing most of the human cDNA library in a year or two. As have many of those sequencing cDNA, including British government scientists and industrial groups, Venter has either filed patents for his sequences or has not yet released them into the public domain.

But Venter's venture may have the cDNA exploration field to itself at present. Few outside investors are willing to put money into further research on cDNA sequences — to learn their function, for example — if they are not sure of owning the product.

However, if bare cDNA sequences turn out not to be patentable — as the White House is now predicting (see below) — anyone would be free to hunt for the biological function of cDNA sequences. Although Venter's group may have a head-start, there are as many as 100,000 genes to explore, more than enough for several companies. Without cDNA patents to muddy the waters, "it's off to the races", says M. James Barrett, president of the Maryland-based Genetic Therapy Inc.

Beyond cDNA research, the best pros-

pects for near-term commercial investment in the genome project appear to be in gene-based diagnostics and those interested in the project as a market. At an estimated \$10 billion a year, the US diagnostic market is huge. But prenatal and parental tests for genetic diseases make up less than one per cent of that, says John Teipel, an analyst with the New-York-based Wilkerson Group.

As the cystic fibrosis (CF) case proved, finding a gene does not create a major market overnight. Three years after the discovery of the gene, prenatal and parental CF screening tests are just appearing and the market is not expected to exceed \$10 million in the near term. "We never felt it was enough to base a company on", says Anne Di Sante, director of the intellectual property office at the University of Michigan, where the gene was discovered. "It's just a gene — not a delivery mechanism for gene therapy."

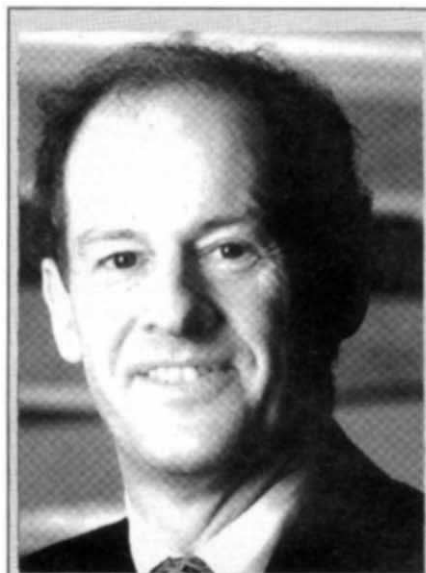
Although the genome project is expected to turn up many other genes associated with genetic disease during its 15-year life, converting those genes into products — using anything from genetic therapy to therapeutics based on expression products — is a long and complicated process, something the biotechnology industry has learned the hard way. "Scientists may be convinced that it's good science, but that doesn't necessarily translate into good business", says Steven Burrill, a biotechnology expert with the San-Francisco-based accounting firm of Ernst & Young.

For the moment, the best bet for investors is companies with products stemming from the genome project itself. Several computer companies, including IBM, are funding genome informatics research in the hope of sharing what is expected to be a major market in processing genome information. Other investors have their eye on the \$3 billion the US government is expected to spend on the project over its lifetime.

Frederic Bourke, the wealthy entrepreneur whose attempt to start a sequencing company earlier this year triggered an international controversy (see *Nature* 355, 483; 1992), says he intends to go ahead with his project despite failing to recruit several prominent researchers. Bourke is considering an investment of \$50–\$60 million to start a company that would operate at first under contracts and grants from the US government and industry.

But he too warns that the genome project is not yet a sure investment, and that it is easy to back the wrong technology or people. Even if scientists agree that the time is ripe for a large sequencing operation, says Bourke, "it's got to be economically viable. I'm not Fort Knox."

Christopher Anderson



Bromley says no to treaty.

US rules out gene patent treaty

Washington. The chief US science official said last week that the United States will abide by its controversial decision to file applications for patents on gene fragments and will not pursue an international treaty on the issue. D. Allan Bromley, science adviser to the president, said that the White House is confident that the US patent office will reject applications for more than 2,000 unknown gene fragments and, therefore, that no policy decision is needed.

Until the patent office makes its ruling, Bromley said, the United States will continue its interim policy of filing patents for fragments of expressed genes, even though their function is unknown. But he expects a ruling much sooner than the usual two or three years the patent office takes for routine applications. In the meantime, Bromley said, the cost to the US National Institutes of Health (NIH) and to companies that are following NIH's lead is "trivial" compared with the overall cost of the project and was not a factor in the administration's decision to file for patents.

Bromley said that a proposal by Britain, France and Japan for an international treaty that would forbid patents on such uncharacterized gene fragments would be like "using an elephant gun to kill a butterfly". Patent offices in Europe and Japan are thought to be even less likely than the US patent office to approve the NIH application, and Bromley believes that the issue can be resolved without the need for a policy confrontation.

C.A.

Delaney's revenge: Court rejects US approach to assessing cancer risks in processed foods

Washington. US risk assessment officials for years have been quietly devising novel interpretations to get around restrictive — and, many researchers argue, scientifically outdated — laws on the use of potential carcinogens in food and water. Last week, however, that regulatory house of cards collapsed.

A federal court in California ruled on 8 July that the US Environmental Protection Agency (EPA) must abide by the letter of the so-called 'Delaney clause' of the 1958 Food, Drug and Cosmetic Act, which prohibits processed foods from containing any chemical found to cause cancer in animals or humans at any concentration. It rejected EPA's argument that it was following the intent of the law to protect consumers from unreasonable risk. As well as reversing decades of regulatory policy and potentially making illegal many of the pesticides now used by US farmers, the ruling strengthens efforts in Congress to overturn the law and to adopt instead a standard that reflects current knowledge of the complex mechanisms of carcinogenicity (see *Nature* 353, 289; 1991).

Improved detection methods have shown that more agricultural chemicals may be present in food after processing than was thought in 1958 to be the case. And more sophisticated testing has shown that some of those chemicals may not be entirely noncarcinogenic. Some, such as saccharine, cause cancer at high doses in mice or rats, but their danger to humans is less well understood. Others, like benomyl, a fungicide used on raisin grapes and tomatoes intended for processing, and phosmet, an

insecticide detected in cottonseed oil, appear to pose definite but slight human risks — of the order of one new cancer in between 10 million and 100 billion people.

EPA studies have shown that at least 67 of about 300 pesticides now used in US agriculture cause cancer in one or more



species of laboratory animal. About 35 of those chemicals are used in processed foods and would be banned under a strict interpretation of the law.

In 1989, realizing that strict adherence to the Delaney clause would make it impossible to approve many of the pesticides now in common use, the EPA adopted a 'negligible risk' policy that permits the use of potentially carcinogenic chemicals if their risk is less than one additional case of cancer per

million people exposed. But last year, the Natural Resources Defense Council challenged that interpretation in court.

"The EPA in effect asks us to approve what it deems to be a more enlightened system than that which Congress established", the appellate judge, Mary Schroeder, wrote in a unanimous opinion. "Revising the existing statutory scheme, however, is neither our function nor the function of the EPA. If there is to be a change, it is for Congress to direct."

EPA officials expected to lose the case. Although the agency could appeal against the decision, it would prefer to see Congress change the law.

Four bills before Congress would do just that. But although all the bills were introduced last year, none has moved ahead because legislators have been awaiting both last week's ruling and a report by the National Academy of Sciences due out later this month.

Now that the California court has passed the baton to Congress, legislators are expected to take up the issue quickly. Twin bills proposed by Henry Waxman (Democrat, California) in the House of Representatives and Ted Kennedy (Democrat, Massachusetts) in the Senate are likely to move the fastest. Those bills, which include children in setting a threshold for any chemical, are more restrictive than some of the other proposals. But in the wake of last week's court decision, both EPA and industry are eager to escape from the Delaney clause, and the Waxman-Kennedy legislation is helped by support from several environmental groups.

Christopher Anderson

One policy for all?

Although the US government would most like to resolve the dilemma of zero-risk posed by the Delaney clause, regulating risks outside agricultural chemicals is also the subject of sharp debate. An interagency committee under the White House Office of Science and Technology Policy has been working for the past year on a set of broad risk-assessment guidelines for all federal agencies, including the Food and Drug Administration (FDA), the Occupational Safety and Health Administration and the Environmental Protection Agency (EPA). The guidelines, which are expected to be issued in the next month or so as a presidential executive order, would create a central review panel for risk decision, perhaps under the direction of the White House.

Industry and environmental groups want to help shape the guidelines, which will be the first major change to US risk-assessment policy in three years. One proposed draft, offered to the White House by Federal Focus, an influential industry-funded group, would encourage regulators to use a variety of risk-assessment methodologies rather than just the so-called linear

modelling now favoured by the EPA.

In linear modelling, regulators assume that a chemical that causes cancer in large doses will cause proportionally fewer (but not zero) cancers at small doses. But some chemicals, such as dioxin, appear to have a 'threshold' concentration below which they may not be carcinogenic. Linear modelling may exaggerate the danger of nonmutagenic carcinogens.

The Federal Focus guidelines also encourage regulators to give equal weight to negative results when making risk decisions. Although many researchers agree that negative results are important, there is concern about lending credibility to flawed studies that may obscure signs of possible carcinogenicity.

Environmental groups are concerned that moving risk decisions away from the science-based agencies such as EPA and FDA to a central group controlled directly by the White House may allow politics to take precedence over science. Ironically, that is just what the Bush administration said was wrong with the Delaney clause when it moved to a new standard in 1989. As proof, environmentalists point to the growing use by the administration of such nonscientific calculations as regulatory-cost-per-life-saved in calculating acceptable risk.

C.A.

UK centres target brain drain

London. Getting British-trained molecular biologists and researchers in related disciplines to return after postdoctoral experience in Europe and the United States is a priority for the directors of two new interdisciplinary centres being set up by the University of London.

"Some people are not coming back because the mid-range jobs have dried up, but now even the best people are not coming back", says Nigel Holder, director of the Developmental Biology Research Centre (DBRC), which was opened last week at King's College London. Holder said that one of the criteria by which he will judge the success of the DBRC is by how many people can be lured home from overseas.

The conditions keeping postdoctoral researchers from returning are also making it difficult for those who have stayed behind. Small, interdisciplinary research centres such as the DBRC and the Laboratory of Molecular Cell Biology (LMCB), due to open at University College London in May 1993, are increasingly being seen as one of the few ways to conduct research of international quality in a university setting.

In opening the DBRC, Sir Mark Richmond, chairman of the Science and Engineering Research Council (SERC), said that too much research in universities is being forced into classical scientific disciplines to match teaching patterns, a situation that is out of touch with the cutting edge of science. "If you are having to talk about interdisciplinary work, that means you have got the definition of the discipline wrong", he said.

Although most of the researchers staffing the DBRC and the LMCB will have departmental posts, the centres themselves are outside the departmental system. The DBRC has 45 research positions spread through ten laboratories, supported by £3.5 million (US\$6.7 million) from research council and charity project grants. Core funding for the building has come from King's College. Two of the group leader posts have been funded externally as five-year contracts with the Medical Research Council (MRC) and the SERC, and both have been filled by the sort of candidates that Holder was hoping to attract.

The LMCB, with plans for between 70 and 90 researchers, also plans to use five-

year contracts under a fellowship scheme to attract postdoctoral students who have stayed abroad. The response so far has been encouraging. "What surprised us was that it was not just people who had just finished postdoc positions who applied, but also people who had been in the US for ten or so years, waiting for a chance to come back," said Colin Hopkins, director of the LMCB. "They appear to be willing to take a cut in status, if not a cut in salary, in order to return to this country."

The LMCB differs from the DBRC in that it is funded under a government scheme for interdisciplinary research centres. It has been built and equipped by the MRC, with the work to be funded by project grants. The LMCB also incorporates an institute whose membership will give like-minded researchers on other campuses access to LMCB facilities and encourage interchange within the molecular biology community. There are also plans to develop a graduate school in molecular biology.

Although developing this sort of research centre is not an option for every British university, there is a feeling that Britain can accommodate several in each field. In molecular biology alone, Hopkins feels that there is room for six or eight centres.

The centres also provide a way for the SERC to exercise greater control over what its researchers do and how they divide their time between teaching and research. The MRC works largely through its own units in universities and hospitals, where its researchers are essentially full-time employees. Similarly, the Agriculture and Food Research Council and the Natural Environment Research Council have institutes devoted to particular areas of research. Most of the recipients of SERC grants, however, work in university department laboratories and face increasing demands to teach.

There are expected to be 300,000 more students in the British university system by 2000, and the bodies governing higher education are devising a plan to accommodate them. The most important changes are likely to involve the idea of extending the length of the teaching day and week and adding a summer term. "If researchers are not protected from these things, when on earth are they going to do research?" asks Richmond.

Starting this autumn, the SERC will become responsible for funding all the direct costs of its research. In the past, those costs were shared with the Universities Funding Council. If the SERC is not convinced that the laboratories of its grant recipients are going to have the full support of the host university, Richmond says, then it may want to consider funding its own institutions. Making his position clear, Richmond said "that is a direction I am absolutely opposed to."

Ian Mundell

Concrete holds key to ESRF's opening

London. Construction engineers at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, are monitoring an expanse of concrete slabs to see if they have remedied a construction problem that could delay the start of experiments and reduce the number of beam lines for researchers.

The error came to light last December, when the floor of the experimental hall was found to be curling. Engineers believe that the concrete slabs making up the floor of the facility were not entirely set, and when exposed to air the tops dried faster than the bottoms.

The problem has been solved by injecting more concrete under the slabs, according to Ruprecht Haensel, ESRF's director general. Only a third of the experimental hall — the minimum required to start the beam-line programme — has so far been treated, leaving it stable if not entirely level. The synchrotron itself has not had the same problems and is already producing 'light'.

Engineers are monitoring the floor to determine the durability of the repairs. The situation is complicated by the fact that the instruments to be installed may be more sensitive than those envisaged when the project was started in 1987.

ESRF officials profess optimism. "The problem has not been negligible, but by no means has it endangered any of the dates for user operations," Haensel says about the project, which technically was six months ahead of schedule when the problem was discovered. But in private there remains talk about tearing up the floor and starting again.

Ian Mundell

Chinese researchers promised more

Beijing. Prominent Chinese researchers last month heard their government's finance minister promise them more money — sometime. At a meeting on 17 June between Wang Bing-Ojan and the Chinese Academy of Sciences, the finance minister said that the government has adopted a policy, to be implemented next year, "of gradually leaning towards more scientific research funds".

The academy has worked hard to convince Wang and other top government officials of the need for additional funding of science. Zhou Guang-Zhao, president of the academy, led Wang on a tour of the academy's chemistry, computation and biophysics institutes, and afterwards Wang spoke of the talented research staff and potential for rapid progress. In the interim, Wang encouraged his audience to look for support from international organizations such as the World Bank.

You Qin Li

Correction

Professor Luc Montagnier says that, contrary to the statement in *Nature* (358, 6; 1992), he was not an adviser to the French government in 1983–85. However, as early as August 1983, Montagnier did warn government officials that production facilities for HIV (then LAV) were urgently needed for blood testing. — Editor, *Nature*.

Multiplicity of migrants

L. L. Katan

An informed debate on the manner in which legislation on migration from food packaging is introduced is badly needed. Until then, scientific resources are being needlessly wasted.

SOMEWHERE above the Atlantic, in a Jumbo jet in 1970, a passenger complained that his whisky tasted odd. The bottle was PVC, and the tainting was traced to vinyl chloride monomer (VCM), which has a distinctive odour. VCM had diffused out of the bottle into the drink; a phenomenon that occurs whenever food or drink contacts a solid surface. The resulting food contaminants are called migrants. At about the same time, VCM was confirmed as a human carcinogen.

In the ensuing uproar, regulatory authorities insisted on a remarkable reduction of VCM levels both in manufacturing plant atmospheres and in PVC used for food contact. The pace of legislation on all types of migration increased. The legislation is precautionary: there is no known case of poisoning from migration since exposure to lead was largely eliminated half a century ago, and none at all from organic migrants. Nevertheless, there are more than a dozen government and trade federation laboratories in Europe alone with substantial migration programmes, as well as many in industry and academic institutions. Substantial — and expensive — toxicological work is also under way.

Migration is just one of the routes by which consumers of food become exposed to micro-contaminants such as fertilizers and pesticides, colourants and preservatives. It is a unique problem in that there is an extra step between source and consumer due to the migration itself. The concentration of a direct additive in the food is known from the formulation; the concentration of environmental contaminants can be estimated, or directly measured. But the concentration, or even the presence of, migrants is usually estimated from their concentration or presence in packaging materials.

So regulations depend on analysis of the packaging material, which reveals an even more severe problem. A plastic is manufactured by polymerization of one or more monomers and/or other starting substances, and solvents and catalysts. Other substances may be used in work-up stages. Because no chemical reaction goes entirely to completion, and polymerization constitutes a series of chain reactions, the plastic contains a spectrum of polymers of different molecular mass, from monomers up to a

high value. It also contains products of side reactions and residues of all the auxiliary materials. Purification stages remove some of these, but inevitably add others.

Frequently, further substances are added to the initial polymer to stabilize it. During conversion to finished packages, some thermal degradation of all the components already present, as well as the polymer itself, is inevitable; degradation is reduced but not eliminated by thermal stabilizers which will, of course, contribute their own residues. Also added during the conversion stage are additives to protect the finished product and enhance its performance. All these are liable to break down during processing — some, such as antioxidants, are intended to do so to fulfill their function. Subsequent reactions between any of the moieties can occur, and all the species can have impurities.

Faced with this situation, the EC Commission, advised by the Scientific Committee for Food, has made a heroic attempt to list every conceivable migrant, and make a toxicity assessment of it. The commission has divided potential migrants into a series of types: monomers and other starting substances, biopolymers, additives, aids to polymerization, and so on. Each of these can be broken down further: aids to polymerization, for example, is divided into 28 subgroups.

Thus there are about 3,000 components on the current versions of the two lists prepared so far, namely monomers and other starting substances, and additives. Most of the entries are categorized as requiring more information before final approval can be given. And for the few hundred monomers that have been approved, with limitations, official analytical methods exist for only a few. It is estimated that there are at least another 3,000 compounds in the queue for consideration. Adding in the other categories mentioned above will easily swell the total to around 10,000.

The time is obviously overdue for some informed common sense. The resources simply do not exist to carry out even modest toxicity testing on more than 10,000 compounds, with development of analytical methods for them, on any reasonable timescale. A totally different approach is needed.

There are various options, not mutu-

ally exclusive. First, in the common law approach, the basic philosophy is not to legislate until a social problem is recognized to exist. Scientifically, this translates into the use of negative lists, which are generally regarded as obsolete. It is also a high-risk strategy, because much damage may be done before the causal factor is identified. Second, the overall migration test could be developed to limit the totality of migration. Only substances presenting a toxic hazard below the limit would be measured. Unfortunately, this option still requires toxicity testing of all migrants that might be present.

A third option is surveillance. Legislative controls would be introduced only when a survey shows contamination to be unacceptable. Fourth is 'flavour of the month', in which the legislating body simply waits for a scare story in the media, and introduces controls for that material. Although obviously unsatisfactory, this happens with disturbing frequency. Fifth is the trial and error approach, in which samples would be analysed from time to time and controls introduced on an *ad hoc* basis.

A sixth approach is QSAR (quantitative structure-activity relationship, in which toxicity is predicted from chemical structure). This can be applied, on a conceivable timescale, to every possible migrant. It could well be useful to establish priorities for testing and legislation, but at this stage is very preliminary. Seventh is a total extractables test, in which all potential migrants are extracted using an aggressive solvent and thus made available for toxicological testing.

The last two options are on the path of least resistance. One is to accept a level of toxicological insignificance, so that testing is carried out only on migrants above a certain limit (with a few possible exceptions, for example carcinogens and cumulative toxins). Validated mathematical models can be used to predict migration to the necessary accuracy. The final option is to do nothing. Recent experience in the former Soviet Union shows that technologically inadequate food packaging causes vastly more misery than has ever been suggested, let alone proven, to occur from migration. □

L. L. Katan is at IFPS Ltd, PO Box 1704, Maldon, Essex CM9 8RX, UK.

Cost of brain disorders

SIR — We were surprised to read (*Nature* 357, 348; 1992) the statement that “there is as yet no certain link between significant progress in understanding neuronal function and its application to any human disease”.

In the half a century since chemical neurotransmission was established as the major mode of communication between neurons, we have seen the introduction of many therapies. A few examples of human diseases where totally new treatments have resulted directly from this fundamental discovery include hypertension, myasthenia gravis, Parkinson's disease, migraine and epilepsy.

Diseases where treatments have been refined as a result of discoveries about neuronal function include depression and schizophrenia, and novel treatments for these and other diseases of the nervous system are likely to follow from current research. Neuroimmunology and molecular neurogenetics are two other areas that have had a significant influence on neurological disease.

Not content with writing something that could be misused by those who are against science, you go on to argue purely in terms of economics that providing more money for research into diseases such as Alzheimer's would lead, if a successful treatment were discovered, to a loss of jobs by those constructing nursing homes and by hospital workers and care-givers.

Are you suggesting that society would choose to condemn sufferers and their relatives to years of misery rather than solve a problem in the management of health care? It would have been much more appropriate for a journal that is sponsoring an international conference on 'The brain in well-being and disease' to support neuroscientists in their endeavours to unravel the workings of the brain so that those who have diseases that affect the very essence of our human nature can have some hope that their terrible suffering will be relieved.

F. E. Bloom (Scripps Research Institute, La Jolla, California, USA); **A. Cowey** (University Department of Experimental Psychology, Oxford, UK); **R. L. M. Faull** (Department of Anatomy, School of Medicine, University of Auckland, New Zealand); **D. G. Grahame-Smith** (University Department of Clinical Pharmacology, Radcliffe Infirmary, Oxford, UK); **J. A. Gray** (Department of Psychology, Institute of Psychiatry, Maudsley Hospital, London, UK); **T. Hökfelt** (Department of Histology, Karolinska Institute, Stockholm, Sweden); **L. L. Iversen** (Neuroscience Research Centre, Merck Sharp & Dohme Research Laboratories, Harlow, UK); **R. R. Lillnas**

(Department of Physiology and Biophysics, NYU Medical Center, New York, USA); **C. D. Marsden** (Institute of Neurology, National Hospital, Queen's Square, London, UK); **D. P. Purpura** (Albert Einstein College of Medicine, Bronx, New York, USA); **M. C. Raff** (Department of Biology, University College, London, UK); **A. D. Smith** (University Department of Pharmacology, Oxford, UK); **S. H. Snyder** (Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA); **P. Somogyi** (MRC Anatomical Neuropharmacology Unit, Oxford University, Oxford, UK)

SIR — I was distressed to read the leading article disparaging the report *The Costs of Disorders of the Brain*. This report argued that the estimated \$401 billion annual costs for brain disorders in the United States justifies a substantial increase in funding for brain research. *Nature* did not contest the validity of these cost estimates, which represent 7.3 per cent of the US gross national product.

Your first argument states that “there is as yet no certain link between significant progress in understanding neuronal function and its application to any human disease”. This statement is anti-scientific and at variance with the substantive advances in our appreciation of brain mechanisms in such common disorders as senile dementia of the Alzheimer's type, schizophrenia and substance abuse. Two decades ago, these were not even recognized as brain diseases, whereas there is now substantial understanding of their molecular and cellular pathology, which has implications for treatment.

Your second argument that curing these disorders will result in loss of employment in the health industry involved in their treatment suggests that we should resume the Cold War to maintain the defence industries. But medical care, like military expenditure, is not a productive use of capital.

The major point of the report was that the investment in brain research is not proportionate to the investment in research on cancer and cardiovascular diseases when the costs to society of the diseases are considered.

This inequity reflects the historical stigma associated with psychiatric, neurological and substance abuse disorders coupled with our past ignorance of brain mechanisms involved in these disorders. The recent rapid advances in neuroscience certainly justify a reappraisal of funding commitments that recognizes the current state of know-

ledge and the costs to society of brain disorders.

Joseph T. Coyle
(President)

Society for Neuroscience,
11 Dupont Circle, NW, Suite 500,
Washington, DC 20036, USA

■ Our case was simply that, among the “more than enough reasons to ask federal agencies to contribute large sums to brain research”, that given (“that the science right now is ripe for it”) is more cogent than that based on the gross cost of disorders of the nervous system. — Editor, *Nature*.

No access

SIR — Whether by intent or serendipity, there was a remarkable and sobering piece of editorial selection in *Nature* for 30 April 1992.

The leading article, “Does science leave room for soul?” (356, 729; 1992) was dispiriting for a variety of reasons, especially because we have all hoped that the anti-intellectualism that it discussed (as well as an imminent fear of the complexity of science) had largely been exorcised by modern education. Clearly and frighteningly, this is far from the case.

Donald Hayes's “The growing inaccessibility of science” (356, 739; 1992) provided part of the explanation. When fewer and fewer scientists can understand one another — and when most of us can come to grips with only a tiny fraction of the Letters to *Nature* — how can the public at large be expected to understand science? And if they cannot understand it, how can it possibly contribute to the elucidation of their world? Is it then any wonder that so many people are turning to religious fundamentalism, astrology and other such duplicitous practices? That development itself would be bad enough, but a few other alarming thoughts follow.

If the public devalues science, how long — as the leading article adumbrated — can we expect that electors will continue to tolerate government funding of our activities? And, indeed, if the mass of detailed science is meagrely accessible to scientists generally, how is new scientific work to be assessed and deceit prevented? Such attacks on the future of science portend a dark age. Are scientists willing to step boldly into the “public culture” and explain our work cogently and lucidly? If we are not, do we deserve a future?

John Carmody

School of Physiology & Pharmacology,
University of New South Wales,
PO Box 1, Kensington,
NSW 2033, Australia

'Blue ostriches' captured

SIR — The 'blue ostriches', a famous portrayal of a supposed rock painting, is very probably a forgery based upon an engraving (see top figure) in Moffat's *Missionary Labours* (1842)¹. The 'copy' of the mysterious painting (middle figure) was made by G. W. Stow, who was born in England in 1822 and migrated to South Africa in 1843². In 1867 he began copying rock paintings and engravings in the Cape Colony and Orange Free State, while preparing a book on the peoples of southern Africa. He died in 1882 before it was completed.

Stow's widow sold the manuscript and rock art copies to Lucy Lloyd, who was studying the Bushmen. In 1904, Lloyd passed the manuscript to the historian G. M. Theal, who published it³. At her death, Lloyd left the copies to Dorothea Bleek who published 74 of them⁴. The rest were published in 1953⁵. The volume edited by Theal described Bushmen disguising themselves as ostriches for the hunt. Opposite page 82 there is what purports to be a copy of a remarkable rock painting (middle figure). It shows five ostriches looking to the right. They are approached from the right by an 'ostrich' with human legs and an arm holding a bow protruding from its chest. Four birds (including the disguised hunter) are in black and white, as in male ostriches, while two presumed females are blue-grey. Bleek republished the 'blue ostriches' in her selection of Stow's copies⁴.

An original 'blue ostriches' rock painting has long been sought. Notes on the cartoon read, "From rocks in the Witte Bergen, Colonial Native Reserve". A map shows a site near Hershel⁴. In the 1905 book³, the caption reads, "From Caves in the Herschel District, Cape Colony . . .". Although Bleek found the sites of 60 other cartoons, an original 'blue ostriches' painting has never been located. It remains "a perpetual challenge"⁶ and "one of the most tantalising mysteries of Bushman art"⁷.

In 1976, P.V.T. acquired a copy of Moffat's *Missionary Labours*¹. Leafing through it he found a virtual mirror image of the 'blue ostriches' painting facing page 64, and immediately suspected that a fraud had been perpetrated. Independently, T.A.D., in 1989,

made the same discovery. In Moffat's engraving, described as a sketch of a Bushman stratagem in hunting, five ostriches form the main group, four looking to the left, while the fifth looks towards the artist. Two have black and white plumage (males), two have grey plumage (females); the fifth is of indeterminate coloration. Approaching from the left is a male 'ostrich' with human lower limbs and, protruding from the chest, a bow held by a hand, the bowstring drawn taut, an arrow tip in position, pointing towards the five ostriches. Apart from mirror imaging, the near identity between Moffat's illustration and the Stow painting is overwhelming (compare the middle and bottom figures).

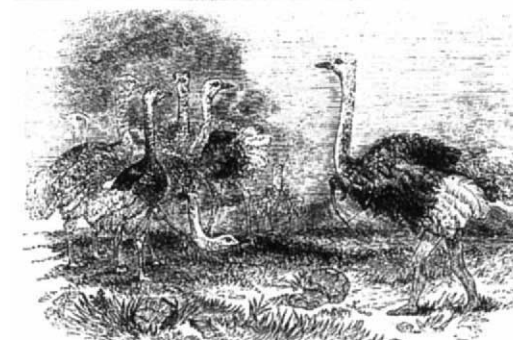
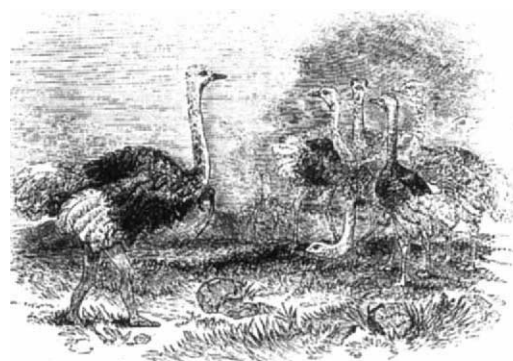
A chance mention revealed P.V.T.'s and T.A.D.'s independent discovery. Meantime, J.D.L.-W. had been studying the Stow archives in the South African Public Library and the South African Museum, to which we are indebted. This is the first joint announcement of what we conclude was a forgery.

Our case rests on two principal points: (1) Stow's picture is nearly identical with a mirror-image of Moffat's previously published wood engraving; (2) an original 'blue ostriches' painting has never been discovered despite searches.

Three supporting points are: (3) the descriptions of the locality are vague (4) Moffat's *Missionary Labours* appeared in 1842; Stow sailed from London in September 1843. It is conceivable and likely that Moffat's book accompanied him. (5) The blue colour of the female ostriches in the cartoon is decidedly unusual in southern African rock art.

Taken together, the evidence leads us to conclude that Stow's 'copy' was fraudulently based upon a mirror image of the engraving in Moffat¹.

There is no evidence that Stow's wife, his three daughters, his friend C. S. Orpen, Lloyd or Bleek had the opportunity, knowledge, artistic skill or motive to have perpetrated the hoax. The evidence from the paper of the cartoon, the style and technique, and the handwriting giving the vague locality points strongly to Stow as the one who faked the 'copy'.



Top, Illustration from R. Moffat's *Missionary Labours and Scenes in Southern Africa* (1842). Middle, G. W. Stow's 'copy' of a supposed rock painting illustrating the Bushman practice of stalking in an ostrich disguise. Bottom, Moffat's illustration reversed.

He had opportunity, knowledge and artistic talent². We find compelling motives. First, the biography² and Stow's letters reveal him as boastful, ambitious and resentful that he had not received sufficient recognition. In Stow's mind, the 'blue ostriches' would assuredly have enhanced his claim to fame. A second possible motive was to provide a 'rock painting' that verified Stow's conviction that rock art constituted "a history of the manners and customs of the Bushmen, as depicted by themselves"².

Our more detailed treatment will appear elsewhere.

P. V. Tobias

Department of Anatomy & Human Biology,

T. A. Dowson

J. D. Lewis-Williams

Department of Archaeology,

University of the Witwatersrand,

Johannesburg, South Africa

1. Moffat, R. *Missionary Labours and Scenes in Southern Africa* (John Snow, London, 1842).

2. Young, R. B. *The Life and Work of George William Stow* (Longmans, Green, London, 1908).

3. Stow, G. W. *The Native Races of South Africa* (ed. Theal, G. M.) (Swan Sonnenschein, London, 1905).

4. Stow, G. W. & Bleek, D. F. *Rock-Paintings in South Africa from Parts of the Eastern Province and Orange Free State* (Methuen, London, 1930).

5. Rosenthal, E. & Goodwin, A. J. H. *Cave Artists of South Africa* (Balkema, Cape Town, 1953).

6. Woodhouse, H. C. *Rock Art* (Purnell, Cape Town, 1978).

7. Woodhouse, H. C. Location mystery (*The Sunday Star Review*, Johannesburg, p. 22, June 29, 1986).

Well-being and productivity

SIR — Three letters have remarked on my Commentary¹ on meat production. Wilson² says: "One problem with using productivity to measure welfare is that the welfare of an animal is a property of an individual, but productivity . . . is usually measured in terms of flock or herd production." But the whole is the sum of its parts, and the welfare of a flock depends upon the welfare of individuals.

Broom³, cited by Wilson², lists the following "indicators of poor welfare": impaired growth, impaired reproduction, body damage, disease, immunosuppression, behaviour anomalies, reduced life expectation, adrenal activity and self-narcotization. The first six of these are pertinent to meat production and I focused on impaired growth, impaired reproduction (egg laying) and disease¹. Broom says "[t]he scientific study of animal welfare should be promoted so that decisions are made on factual rather than emotional grounds", which is in accordance with my commentary¹.

Wilson alleges that "a mortality of 7 per cent is not unusual in rearing veal calves to six months". I doubt that producers would be willing to accept 7 per cent mortality. Wilson next discusses measurement of egg productivity in poultry by egg mass output or egg size. An economic decision would depend upon the same price per egg versus a premium for large eggs. He then states that "laying hens kept in battery cages often have poor feather cover and poor skin condition . . .". In such cases, alleviation should be obtainable by improved husbandry. Keeping hens (and other chickens) indoors protects them from predators, bad weather and insanitary conditions. Next, he says that "foxes kept in cages usually have good fur and skin condition, but often have low reproductive rates". This seems off the subject. Foxes are not used for meat, and they are wild animals, which might explain their low reproductive rates in captivity. Wilson then alleges that I have given credence to folklore. This is incorrect: I based my arguments on impaired production and disease prevention. Production values have been an important adjunct to aiding animal welfare. However, some of his contentions are based on hearsay, including conversations with farmers.

Schönthal's letter⁴ is unsupported by references. He says I should ask myself why "chickens get their beaks cut". De-beaking, really beak-trimming, is by removal of the nonvascular tip of the beak in a manner similar to paring fingernails. This protects other birds from attack; however, it is not widely used in broiler

and fryer production because it is not necessary. One company in California hatches 3.5 million chicks weekly and does not use it for broilers and fryers, but breeding males are debeaked. In Georgia, a current practice is simply to dim the lights. I used this procedure successfully years ago. Schönthal misquotes me by saying that I referred to "these animals jammed in cages". Actually, I said that "overcrowding reduces weight gain". Schönthal wishes to compare the situation of crowded animals to "Hitler's concentration camps". This is an emotional and inaccurate analogy. Hitler's victims were starved to emaciation and death. Animals used in meat production are well fed and nurtured. Chickens often tend to crowd voluntarily at one end of their quarters. They do not usually try to escape as long as there is abundance of food.

Schönthal makes a common mistake in claiming that farm animals today are "much more prone to diseases because of the conditions in which they are raised". On the contrary, they are less prone. Animals in extensive, as opposed to intensive, production systems are exposed to greater environmental extremes (see above), to increased "exposure to internal parasites" and have an associated higher disease susceptibility⁵. The control of common poultry diseases, also of predators, is best achieved under confinement rather than "out on the range".

Rowan⁶ pleads for the development of "confinement systems that retain most of the advantages for farmers and the animals while eliminating most of the adverse effects". I agree: my point was that too-close confinement systems are usually counterproductive, and that this has led to their improvement.

Thomas H. Jukes

University of California,
Department of Integrative Biology,
Berkeley, California 94720, USA

1. Jukes, T. H. *Nature* **355**, 582 (1992).
2. Wilson, A. D. *Nature* **356**, 556 (1992).
3. Broom, D. M. *J. Anim. Sci.* **69**, 4167–4175 (1991).
4. Schönthal, A. *Nature* **356**, 556 (1992).
5. Hays, V. W. *Agricultural Uses of Antibiotics*, 7 (American Chemical Society, Washington DC, 1986).
6. Rowan, A. N. *Nature* **356**, 556 (1992).

UK science

SIR — Terence Kealey asserts that only an absolute growth in British science is sufficient to meet national needs and that the relative decline of science is both inevitable and desirable (*Nature* **358**, 272; 1992). These are erroneous beliefs and, as guides for science policy, misleading. With only an absolute

expansion, national living standards decline and unemployment rises. This would hardly qualify as meeting national needs.

The explanation is fairly simple. As economics mature, resource costs rise, particularly labour. Cheaper imports from newly industrializing nations find a ready market, as labour costs in these poorer countries are held stable at low levels by an abundant workforce and high unemployment. The competition from lower wage rates can be offset in mature economies by increasing labour productivity and worker skills through the application of technological innovations. To maintain living standards and employment levels, higher productivity skills must improve job performance for those just entering the workforce and also for workers displaced by cheaper production from poorer countries. Unless improvements encompass both groups, wage rates can be expected to fall and production decline. The British economy has a larger share of workers in trade-sensitive industries than more insular economies and therefore requires a larger effort.

The gains from the application of technology can be seen from US experience. The observed increase in the differential earned in the 1980s by the more highly educated portion of the US workforce came from the use of new technologies such as computers, communication, photocopying and other capital equipment incorporating high technology (J. Bound & G. Johnson *Am. econ. Rev.* **82**, 388–389; June 1992).

There is no reason to believe that the source of improved labour-force skills is limited by national frontiers. It is also questionable whether citations of the scientific literature can serve as a reasonable proxy for those skill-enhancing factors. Whatever the nature of productivity improvements, it is their adaptation that is of primary importance. Presumably, adaptation is more likely to take place in the country of origin and publication of the science. An example to the contrary, however, is the application by Japanese industry of science and technology breakthroughs originating in the United States.

What is not tenable is the equanimous acceptance that a minimal expansion in science is sufficient to meet national needs. Such a policy leads to falling wages and rising unemployment. Surely the factors that make a decline inevitable are more likely to be found in sloth and ministerial neglect. There are many examples in history to show that both conditions are reversible.

Edward J. Krowitz

2415 North Dickerson Street,
Arlington,
Virginia 22207, USA

How to publish the unpalatable?

Professor Philippe Rushton from the University of Western Ontario is skating on thin ice in seeking to use an analysis of the cranial capacity of US Army soldiers to support his views on racial differences in IQ.

How should a journal, this one perhaps, respond when asked to publish material that contradicts generally accepted views, that the Earth is a rough approximation to an ellipsoid for example? Or that Einstein's relativity is broadly correct (if not widely understood)? The proper response is that a person claiming to be able to stand generally accepted views on their head has an urgent claim on public attention, but that he or she must bring to the task evidence that is overwhelmingly compelling. There are only a few occasions when a hypothesis under attack is so clearly and elegantly formulated that it will collapse in the face of a single falsification. It took a small army of people half a century to vindicate Wegener's view that continents drift. Most people (but not all iconoclasts) appreciate that painful truth.

Should journals respond differently when what the iconoclasts advocate is unpalatable for other reasons, merely by being politically incorrect, for example? (Peter Duesberg's belief that human immunodeficiency virus is not the cause of AIDS qualifies under both headings.) Strictly speaking, unpalatability of that kind is irrelevant to the endless search for what the truth may be, and thus should be irrelevant to the question of publication. But that is only part of the reality. If, for example, a psychologist claims to have proof that personal liberty is so great a psychological embarrassment for most human beings that the classical model of democracy is insupportable, journals should again be eager to publish, but should insist that because the general conclusions are so capable of being misunderstood, the evidence must again be unusually compelling.

So what if an author claims evidence to show that Blacks (people with black skins of African origin) are more intelligent than Whites (called Caucasians), as measured by IQ? Evidently the conclusion would be of great importance, and should be published quickly. But all journals know that the intellectual and other behavioural attributes of race have been contentious, even divisive, issues since the beginning of experimental psychology. (Before that, people took them for granted.) And such would be the general excitement and the scope for misunderstanding occasioned by publication that the requirement that the evidence should be extraordinarily compelling would again apply. To fail in that regard would be unseemly.

What follows is mostly about race and IQ or, rather, about race and cranial capacity, taking the latter to be a proxy for IQ. Next week in Brussels, the International Congress

of Psychology is to hear a talk from Professor J. Philippe Rushton, from the University of Western Ontario, who has put out an advance summary of what he plans to say. Briefly, he claims that men have a greater cranial capacity than women, that educated people have larger brains than "those lower in achievement" and that cranial capacity can be ranged in decreasing order from Asian Americans to White Americans to Black Americans.

Rushton is a controversy in himself. For several years, he has been concerned with racial differences in IQ and related matters. Inevitably, his interests, which are politically incorrect, have landed him in trouble. There have been student demonstrations and protests from his colleagues at Western Ontario. For a time, he was suspended from his faculty position. Rushton has diligently fought his corner, on the proper grounds that for a person to be pilloried for the unpopularity of his views is a threat to academic freedom. So it is, or would be. But on the evidence of documents now distributed, there is more than that to say.

First, as will be freely acknowledged, the database on which Rushton's study is based is huge and without precedent. It derives from the US Army's habit of collecting anthropometric data to help with specifying what sizes of helmets (and other equipment) it should manufacture. Data on 9,000 Army personnel were collected in 1988, from which Rushton extracted the records of 6,325 soldiers of all races. (He says he "winnowed the pool" to exclude ambiguity, including only those who had defined themselves as Asian, Black or White.)

Cranial capacity is calculated from the length, breadth and height of the skull by a formula first published by Lee and Pearson in 1901. The analysis is necessarily complicated: there are 12 categories of subjects (3 racial groups, 2 genders and 2 ranks — officers and enlisted soldiers separately), while the effects of weight and stature (correlated positively with derived cranial capacity) must be allowed for. The essence of the result is that the cranial capacity of Asian-American, Caucasian-American and African-American soldiers are 1,403 cm³, 1,361 cm³ and 1,346 cm³ respectively.

Rushton remarks on his data that, "As one reviewer noted, they may be the most valuable set of data in this whole literature." But the self-congratulation is misplaced. Certainly it is a remarkable dataset, but it cannot sustain the generality of the conclusions that Rushton draws from it even in the limited matter of the variations of cranial capacity.

Rushton says that his sample is a "stratified random sample", but that can be true of the US Army's population only, not of the US population as a whole.

Worse, there is every reason to expect that it will be a biased sample: the reasons why different genders or racial groups join the US Army are notoriously different. It could be that Whites enlist to make careers for themselves, and that the Blacks who join are those who cannot find other jobs. If that is the case and if Rushton is also correct in asserting a correlation between cranial capacity and IQ, the differences of cranial capacity he finds are just those he should have found if the US Army consisted of differently biased samples from the two racial groups. Only a little imagination is required to think of other causes of bias.

In other words, despite the care with which the influence of stature may have been controlled, the study entirely neglects bias in the composition of the sample determined by cranial capacity or by the IQ for which Rushton takes it to be a proxy. The data, whether "the most valuable . . . in this whole literature" or not, may say something about the US Army, but can say nothing about Whites and Blacks in general.

This is an elementary error to find in an article with such an important conclusion. People seduced by the wonderful complexities of the analysis of variance (ANOVAR is the word) are often forgiven by their colleagues for not searching, as they should, for causes of bias in the data to which they apply their tricks. On this occasion the oversight will be less readily overlooked.

This elementary error in Rushton's argument will allow many readers to avoid the difficulty that follows of grappling with his speculations about the possible causes of the most marked observation: that women's brains seem to be 100 cm³ smaller (on the average) than men's. Are women's neurons more tightly packed, or do they have less myelin? When so little is known of the working of the brain, let alone of the connection between neuronal activity and measured IQ, these speculations have a distinctly old-fashioned ring.

The more serious complaint is that an argument putting forward a politically incorrect conclusion in a manner likely to be widely and generally misinterpreted has not been required to meet the test that the proof should be especially compelling. It is surprising that Rushton, with his recent experience, has not tumbled to the need off his own bat.

John Maddox

A tool for transgenesis

Cheryll Tickle

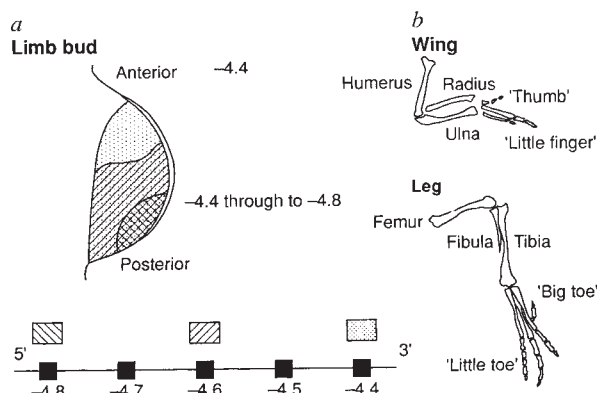
A FUNDAMENTAL problem in embryology — how structures arise in the appropriate places¹ — is exemplified by the development of the limb. How is it that the little finger develops at one edge of the hand and the thumb at the other? The report by Morgan and colleagues on page 236 of this issue² seems to provide some answers and points to a causal relationship between expression of homeobox-containing genes of the Hox-4 complex and finger (or toe) identity.

The initial clue that Hox gene expression could encode the positional address of cells in the limb bud came from examining the expression patterns of gene members of part of the Hox-4 complex³. The domains of expression of gene members *Hox-4.4* through to *Hox-4.8* (at the 5' end of the complex) overlap at one edge of the limb bud (see figure), and cells at different positions across the limb bud express different combinations of these genes. Cells in the posterior part of the bud that will give rise to posterior structures such as a 'little finger' express all the genes, whereas anterior cells that will give rise to the anterior 'thumb' express only *Hox-4.4*.

At this stage, enter the chicken embryo. The limb buds of chick embryos can be manipulated by the cut-and-paste type of experiment. When the polarizing region, a signalling region which is located at the posterior edge of the bud, is grafted to the opposite edge (the anterior) of a second bud, an additional set of digits develops in mirror-image symmetry with the normal set of digits⁴. The same effect can also be produced by local application of retinoic acid, which is a candidate for the polarizing region signal. Last year, two groups showed that both polarizing region grafts and retinoic acid application lead to mirror-image patterns of Hox-4 gene expression that herald the subsequent mirror-image pattern of digits^{5,6}. The behaviour of the genes in the Hox-4 complex also provides a molecular explanation for several embryological observations. For example, because activation can proceed in only one direction along the complex, this explained why manipulations can convert anterior structures into posterior ones, but never posterior into anterior. Nevertheless, despite these encouraging consistencies, it was still possible that expression of Hox-4 genes was one of the consequences of the change in cell

fate rather than a cause of it.

The work by Morgan *et al.*² convincingly bridges this conceptual gap. The experiments involve a new technique for inserting genes into the cells of chicken embryos (more of which later), enabling the authors to alter the expression domain of one of the Hox-4 genes, *Hox-4.6*, in a limb bud. *Hox-4.6* is normally expressed throughout a substantial part of the bud but not in anterior cells which give rise to the 'thumb' or 'big toe'. If



a, Hox-4 gene expression in the early limb bud and, b, the pattern of cartilage elements in the chick wing and leg. Cells in posterior regions of the bud express *Hox-4.4* through to *Hox-4.8*, whereas cells in anterior regions express only *Hox-4.4*. The expression domains of three of the genes are illustrated.

Hox-4 gene expression determines the fate of cells in the limb bud, then altering the spatial pattern of expression of a member gene should alter the identity of digits in a predictable manner.

The most dramatic result is seen in the leg. The anterior toe ('big toe') of a chicken foot is shorter than the rest, has only two cartilage elements (see figure) and is normally reflected so that it points backwards. In legs that develop from infected buds in which all the cells express *Hox-4.6*, this toe is longer, has three cartilage elements and is not reflected. In other words it looks the same as the adjacent 'index' toe. Changing the combination of Hox-4 gene expression in anterior cells to a more posterior one makes the phenotype of the digit more posterior.

In the wing, alteration of Hox-4 gene expression leads to the development of an additional 'thumb' rather than a change in digit phenotype. This result is rather more difficult to explain without making other assumptions. If the region of the wing that will form the anterior digit is, in effect, enlarged by changing the pattern of Hox-4 expression, it is not clear why a discrete additional digit is obtained rather than one very fat one.

The same phenomenon is also seen in the polydactylous chicken mutant, *talpid* (*ta*³). Here, the expression of Hox-4 genes across the limb bud is uniform but up to ten morphologically similar digits develop⁷. The mechanism that controls digit number may not be the same as that controlling digit identity.

So far only the digits have been considered — what of other parts of the limb pattern? It is well established that the cartilage elements of the limb are laid down in sequence starting with femur/humerus, followed by a pair of bones and ending with the digits⁸. Because the expression domains of Hox-4 genes are established soon after the limb bud develops and are maintained as the bud grows out², it looks as though the same combination of Hox-4 gene expression will 'encode', say, both tibia and 'big toe'.

The results of Morgan *et al.* may be consistent with this view. In legs that develop from infected buds, the shank is very short. One possibility is that two fibulae, which are splint-like bones, have developed. In the wing, evidence for conversion of radius into ulna would be more difficult to recognize. The description of the expression domains of Hox-1 paralogues⁹ suggests that these may act in concert with Hox-4 genes to give specific codes that distinguish, for example, tibia from 'big toe'.

Until now, the only tools available to test the causal link between gene expression and pattern in vertebrates have been transgenic mice. The technology reported here² is an exciting innovation. A replication-competent retrovirus carrying the inserted gene was used to infect early chick embryos and the virus spread locally as the embryo developed. There are a number of potential advantages to this type of 'transgenic' chicken approach. One is that not every cell in the embryo needs to carry the inserted gene; this overcomes a particular problem of dealing with a gene that has fundamental importance at several stages in development, because inappropriate expression in every cell

1. Wolpert, L. *J. theor. Biol.* **25**, 1–47 (1969).
2. Morgan, B. A., Izpisua-Belmonte, J.-C., Duboule, D. & Tabin, C. *J. Nature* **358**, 236–239 (1992).
3. Dollé, P. *et al.* *Nature* **342**, 767–772 (1989).
4. Saunders, J. W. & Gasseling, M. T. in *Epithelial-Mesenchymal Interactions* (eds Fleischmajer, R. & Billingham, R. F.) 78–97 (Williams & Wilkins, Baltimore, 1968).
5. Izpisua-Belmonte, J.-C., Tickle, C., Dollé, P., Wolpert, L. & Duboule, D. *Nature* **350**, 585–589 (1991).
6. Nohno, T. *et al.* *Cell* **64**, 1197–1205 (1991).
7. Izpisua-Belmonte, J.-C., Ede, D. A., Tickle, C. & Duboule, D. *Development* **114**, 959–963 (1992).
8. Summerbell, D., Lewis, J. & Wolpert, L. *Nature* **244**, 492–496 (1973).
9. Yokouchi, Y., Sasaki, H. & Kuroiwa, A. *Nature* **353**, 443–445 (1991).
10. Davidson, D., Crawley, A., Hill, R. E. & Tickle, C. *Nature* **352**, 429–431 (1991).

could be lethal to the embryo. Another advantage is that genetic manipulation could be combined readily with experimental manipulation.

The strategy should be widely applicable. For instance, there is the prospect of being able to test the specific roles of other genes expressed in the limb. Hox-7 and Hox-8 seem to be players in the continuous interactions that maintain a region of undifferentiated cells at the tip of the bud from which the pattern is laid down¹⁰. What would happen if these

genes were expressed throughout the limb? Would the bud remain undifferentiated and just get longer and longer? Will it be possible to analyse the role of growth factors? And, of course, it should be possible to apply the same technology to other parts of the embryo. □

Cheryll Tickle is in the Department of Anatomy and Developmental Biology, University College and Middlesex School of Medicine, Cleveland Street, London W1P 6DB, UK

HIGH-ENERGY ASTRONOMY

The big accelerator in the sky

E. S. Phinney

THE positron, the antimatter counterpart of the electron, was discovered on 2 August 1932, dramatically and unexpectedly confirming Dirac's new quantum theory of the electron. The positron in question was created by a cosmic ray which happened to end its wanderings around the Galaxy in the Caltech cloud chamber of Carl Anderson. Astrophysical objects, which have been accelerating particles for much longer and to much higher energies than have terrestrial physicists, can be astonishingly efficient. Near the centre of our Galaxy, some object (or objects) can, when it chooses, create 10^{10} tonnes of positrons per second, with very little waste heat. The positrons, of rest-mass energy 511 keV, annihilate with the electrons they meet, often releasing a pair of 511-keV γ -rays. The 20-year search for the source of these positrons takes an unexpected turn on page 215 of this issue¹.

Radiation

The 511-keV positron annihilation radiation from the inner portion of our Galaxy has been observed with many balloon and satellite-borne instruments since the 1970s (for a review, see ref. 2). Instruments with good energy resolution show that there is a narrow annihilation line centred at 511 keV. Unfortunately these instruments have all had poor angular resolution (15–130°), and because of the heterogeneity of the instruments it has been difficult to make sense of the observations. But it is now generally accepted that the narrow line has both an extended (over 30°), steady component, and a presumably point-like component which varies on timescales of several months. The extended component can be understood as being due to the annihilation in the interstellar medium of positrons released by the radioactive decay of unstable isotopes created in novae and supernovae.

The variable component of the narrow

511-keV line must be more exotic. The fact that it varies in less than a year requires that the positrons slow down and annihilate in less than a year. This sets a lower limit to the hydrogen density n_H in the annihilation region. If the positrons are injected with kinetic energies comparable with their rest-mass energy, the bottleneck is in their initial slowing down, requiring $n_H > 10^5 \text{ cm}^{-3}$. If they are injected at lower energies, they slow down more rapidly, but in an ionized medium with $T \sim 10^4 \text{ K}$, radiative recombination with electrons (or possibly capture on dust) becomes the bottleneck. The density required to recombine in a year is again about 10^5 cm^{-3} . The width of the line, less than 3 keV, requires only that the medium be cooler than 10^5 K , but does not (as once thought) rule out annihilation in a neutral medium².

The poor angular resolution of the detectors gave theorists wide latitude to speculate on the source of the positrons. They rapidly divided into two camps: the exotics and the bores. The exotics argued that such a unique source must be a unique object in a unique place: namely a supermassive black hole at the centre of the Galaxy. The bores pointed out that the total luminosity in photons with energies above 100 keV from the inner 15° of the Galaxy was only a few times greater than the positron annihilation luminosity. This required the positron generator to be very efficient. As positrons are best produced by photon-photon collisions, high efficiency means that a γ -ray created in the source must have a high probability of colliding with another γ -ray. At fixed source luminosity, the density of target γ -rays scales inversely with the square of the source size. Thus high efficiency requires a small source, in this case less than 10^8 cm . This is much too small for a supermassive black hole, but, as the bores pointed out, is natural for an accretion

disk around a neutron star or stellar-mass black hole. The luminosity would be right too. The observed correlation between the annihilation line flux and the flux in higher-energy γ -rays seems to confirm such a simple quasi-isotropic photo-production model.

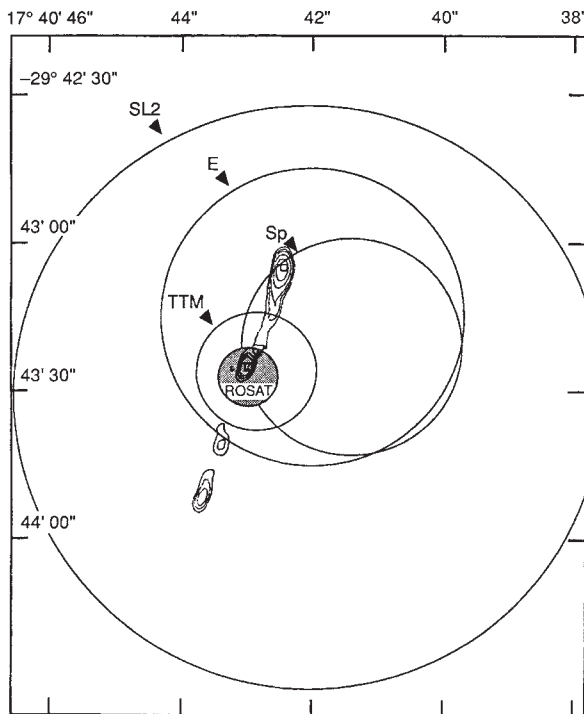
So matters stood in the mid-1980s, when there arrived on the scene γ -ray telescopes with coded mask apertures, giving angular resolutions some hundred times higher than did the earlier instruments. To the dismay of the exotics, these showed that there is no hard-photon or annihilation-line source at the centre of the Galaxy. And to the joy of the bores, they showed that there is a very luminous source of high-energy γ -rays 1° from the Galactic Centre, with the delightful name 1E1740.7–2942. This highly variable source has now been observed more than a dozen times. On 13 October 1990 it showed itself worthy of this intense scrutiny. On that day, its spectrum developed³ what appeared to be a broad line, centred at 410 keV, with width 180 keV and flux $10^{-2} \text{ cm}^{-2} \text{ s}^{-1}$.

Accretion

The obvious interpretation⁴ is that this is the intermittent source of the positrons, and that some of them annihilate in the hot gas of the accretion disk, producing the broad line, and the rest escape and annihilate in the molecular cloud against which the source is projected (see Fig. 3 on page 217 of this issue¹). The broad-line flux on that day was ten times larger than that of the narrow annihilation line at its 1978 maximum. The narrow line seems to be 'on' about half the time, so 1E1740.7–2942 could be in its positron-producing state only 5 per cent of the time (just consistent with being caught once in a dozen looks), and still avoid straining photoproduction efficiencies.

Two possible radio counterparts to 1E1740.7–2942 were discovered byulkarni, Prince and others in 1989⁵ at 5 GHz. Mapping at 1.4 GHz showed⁶ that one of these had a steep spectrum, and the other flat. Mirabel *et al.* on page 215 of this issue¹ have followed up these discoveries by observing the source with the Very Large Array radio telescope in New Mexico in a lower resolution mode, more sensitive to features of low surface brightness. It turns out that the flat-spectrum source is the core, and the steep-spectrum source the brighter lobe of a triple radio source. They suggest that these track twin jets of positrons from 1E1740.7–2942 out to their annihilation sites in the superposed molecular cloud.

This radio source looks exactly like an ordinary extragalactic radio source, with steep-spectrum lobes and a flat-spectrum core. Could it be just that? As Mirabel *et al.* correctly point out, the probability



Error circles of the Great Annihilator 1E1740.7-2942, as observed by various X-ray instruments (SL2, Spacelab 2; E, Einstein; Sp, Spartan; TTM, Mir) and the new radio map¹ of the source. The smallest circle is from a Rosat HRI observation (T. Prince, personal communication and ref. 7), whose error circle is dominated by uncertainties in the alignment of the boresight, still being calibrated. The position and diameter of the circle are based on the provisional calibration as of June 1992. The error circle for the broad annihilation line³ is about twice the diameter of the figure. The error circle for the narrow annihilation line² is (on the scale of the figure) about 27 metres in diameter.

that a random extragalactic radio source more than half that bright would have its core within the error circle of what was the best X-ray position (TTM in the figure) is only 10^{-3} . And the source has passed its first test, as it also lies within the new Rosat⁷ error circle, which is half as big, halving that probability.

But most astronomers are loath to accept these low probabilities as proof that the source is not a chance coincidence. The brightest X-ray source in the sky, Scorpius X-1, engaged theorists for 20 years because of a perfect association with a pair of radio lobes not unlike those reported by Mirabel *et al.*. Despite the 10^{-5} probability⁸ of a chance alignment, proper-motion studies last year⁹ showed that Sco X-1's lobes are unrelated extragalactic radio galaxies.

A pessimist would observe that the source of Mirabel *et al.* is 100 times more likely to be a chance association with 1E1740.7-2942 than Sco X-1's radio lobes were with it. An optimist would argue that surely God would not be so cruel as to make both the brightest X-ray source and the brightest positron source in the sky coincide with extragalactic radio sources. The cruellest joke would

be if the X- and γ -ray source were also extragalactic — a quasar or a BL-Lac object. The γ -ray luminosity and variability would not be very unusual, but the radio source would be — in such sources, the radio core is usually much brighter than the lobes.

If the radio source really is in our Galaxy and is truly associated with 1E1740.7-2942, what do we learn? This would, with the unique precessing jets of SS433, be the only evidence of highly collimated outflow from an X-ray source in our Galaxy (Cygnus X-3 has weakly collimated episodes of ejection at relativistic speeds¹⁰). Protostars, with similar accretion disks, but much shallower potential wells than neutron stars and black holes, have hitherto been the only galactic sources with (nonrelativistic) jets as beautifully collimated¹¹ as those produced by quasars and radio galaxies.

The minimum pressure in the radio source lobes would be 10^{-8} dyn cm⁻². If the actual pressure is somewhat greater than the minimum, the lobe velocities

would exceed 10 km s⁻¹, and proper motion might be detectable, as might unusual molecular emission lines doubled by the expanding lobe shocks propagating into the molecular cloud. The column density to the X-ray source is of the order of 10^{23} cm⁻², somewhat less than the length of the radio source times the minimum density for annihilation of the variable narrow line, and less than the inferred column density of the superposed molecular cloud. This suggests that the source is near the surface of the cloud. Compton reflection of the γ -rays and the 170 keV backscatter peak of the annihilation line might thus be observable¹². The X-ray column suggests a dust extinction of 5.8 magnitudes (200-fold) at 2.2 μ m wavelength (K-band). There is no star at the position of the core of the radio source brighter than $K = 16.5$ (G. Neugebauer, personal communication), which would rule out accretion from an O star or blue supergiant (as in the case of Cygnus X-1 and LMC X-1), but not from B stars (as for LMC X-3) or low-mass stars.

Such stars do not produce enough ionizing photons to photoionize the annihilation region. Nor would

1E1740.7-2942 (assuming a standard accretion disk). It would create only a tiny (10^{16} cm) highly ionized zone, surrounded by a huge zone kept weakly ionized (a fraction of a percent) by the penetrating keV photons. Charge-exchange reactions dominate the ionization balance in this zone, so unusual molecules might form, and cooling would be by dust thermal emission, and the Si II 35- μ m and O I 63- μ m fine-structure lines. The narrow positron line would come from the boundary layer of the radio lobes, where positrons mixed with the cloud medium.

Triumph

In this interpretation, the bores would have triumphed. The Galaxy contains dozens of neutron stars and stellar-mass black holes accreting from companion stars. Many have spectra as hard as 1E1740.7-2942's, and might have similar positron jets. But only if located in a dense molecular cloud core would their positrons annihilate rapidly and their jets have high enough surface brightness to be detected. Near the Sun and in the galactic bulge, such dense clouds occupy a much smaller fraction of the volume than in the plane close to the Galactic Centre, so the presence of the Galaxy's one Great Annihilator close to (but not at) the Galactic Centre is not unreasonable.

Radio searches for positronium recombination lines, previously concentrated at the Galactic Centre¹³, might fruitfully be attempted near the lobes of the radio source. And someone should fly a coded-mask γ -ray telescope with a germanium detector, to determine whether 1E1740.7-2942 is indeed the source of the variable narrow annihilation line. □

E. S. Phinney is in the Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, California 91125, USA.

- Mirabel, I. F., Rodríguez, L. F., Cordier, B., Paul, J. & Lebrun, F. *Nature* **358**, 215-217 (1992).
- Ramaty, R. & Lingenfelter, R. E. in *Gamma-Ray Line Astrophysics* (eds Durouchoux, P. & Prantzos, N.) 67-86 (Am. Inst. Phys., New York, 1991).
- Sunyaev, R. *et al.* *Astrophys. J.* **383**, L49-L52 (1991).
- Ramaty, R., Leventhal, M., Chan, K. W. & Lingenfelter, R. E. *Astrophys. J.* **392**, L63-L66 (1992).
- Prince, T. A., Skinner, G. K., Kulkarni, S. R., Matthews, K. & Neugebauer, G. *IAU Circ.* No. 5252 (1991).
- Gray, A. D., Cram, L. E. & Ekers, R. D. *Mon. Not. R. astr. Soc.* **256**, 277-280 (1992).
- Prince, T. *et al.* *Bull. Am. astr. Soc.* **23**, 1392 (1991).
- Geldzahler, B. J., Fomalont, E. B., Hilldrup, K. & Corey, B. E. *Astr. J.* **86**, 1036-1041 (1981).
- Fomalont, E. B. & Geldzahler, B. J. *Astrophys. J.* **383**, 289-294 (1991).
- Molnar, L. A., Reid, M. J. & Grindlay, J. E. *Astrophys. J.* **331**, 494-508 (1988).
- Reipurth, B. in *Physics of Star Formation and Early Stellar Evolution* (eds Lada, C. & Kylafis, N.) 497-538 (Kluwer, Dordrecht, 1991).
- Bjeldsten, L. & Zurek, W. H. *Astrophys. J.* **329**, 212-224 (1988).
- Anantharamaiah, K. R. *et al.* in *The Center of the Galaxy* IAU Symp. 136 (ed. Morris, M.) 607-616 (Kluwer, Dordrecht, 1989).

Cytosolic chaperonin confirmed

John Ellis

EXCITEMENT in science is sparked when the accepted view of a process as basic as protein folding is challenged by a different view that makes predictions confirmed by subsequent experiment. Such confirmation is provided by two papers on pages 245 and 249 of this issue^{1,2}, and by one in *Cell*³ — they report that a protein found in eukaryotic cells functions as a molecular chaperone in assisting protein folding in the cytosolic compartment.

Textbooks tell us that nascent polypeptide chains fold spontaneously into conformations of lower free energy which possess the specificities required for their biological functions. This idea of protein self-assembly was challenged five years ago by an alternative view which postulates that protein folding, as well as other aspects of protein assembly *in vivo*, involves interactions with pre-existing proteins that act as molecular chaperones⁴. These interactions are transient and often require ATP hydrolysis, and their job is to inhibit incorrect assembly pathways which might otherwise produce non-functional structures. This new view of protein folding is most succinctly described as assisted self-assembly, because known molecular chaperones do not convey the steric information essential for assembly; much recent evidence supports this view⁵⁻⁷.

Two groups

Two different groups of molecular chaperone are believed to cooperate sequentially in assisting protein folding, the heat-shock protein 70 (hsp70) group and associated proteins, and the chaperonins⁸. The chaperonins are a group of proteins, highly related by sequence, found in all eubacteria, mitochondria and plastids that have been examined (they sport a variety of other names, for instance GroE proteins and hsp60 proteins). There are two related types of chaperonin, chaperonin 60 and chaperonin 10, the numbers indicating the relative molecular mass of their subunits divided by a thousand. Unlike the hsp70 proteins, which are found in every cellular compartment where protein folding occurs, the chaperonins are confined to eubacteria, mitochondria and plastids, all of which are thought to be related in their evolutionary origins by endosymbiosis. But what about the eukaryotic cytosol, where most protein folding occurs, and which is held to originate from the archaeobacteria?

Two independent studies have used slight sequence similarities to predict that the mouse t-complex polypeptide-1

(TCP-1 protein) might be a cytosolic chaperonin^{9,10}, a view encouraged by a recent report that the archaeobacterium *Sulfolobus* contains a protein (thermophilic factor or TF55) highly related in sequence to TCP-1¹¹. The TF55 protein has ATPase activity and binds unfolded polypeptides, properties characteristic of chaperonins, but there is no direct evidence that TF55 is involved in protein folding. The new papers¹⁻³ provide evidence that TCP-1 does indeed assist the folding of several cytosolic proteins, but intriguingly it shows structural features that distinguish it from the other chaperonins.

The TCP-1 protein was first identified in mice, where it is expressed constitutively in almost every type of cell, but homologues are known from other mammals, yeast, *Drosophila* and *Pisum*. Contrary to a previous report that TCP-1 is membrane-bound, the new data² show that it occurs in the cytosol of murine and human cells as an oligomeric particle of relative molecular mass (M_r) 800–950K which contains, besides TCP-1 (subunit M_r 60K), four to six other unidentified proteins (subunit M_r 53–72K) and variable amounts of two hsp70 proteins.

Electron microscopy indicates that the particle consists of two stacked rings of eight or nine subunits each, with a large central cavity; these particles appear as striated rectangles in side view. This structure resembles that of the TF55 particle of *Sulfolobus*, which occurs as a homo-oligomeric complex of two stacked nine-membered rings, each subunit showing 36–40% amino-acid sequence identity to the TCP-1 protein¹¹. The chaperonin 60 proteins also occur as oligomers of two stacked rings with a central cavity and a striated side view, but here each ring consists of seven identical 58K subunits (except for the plastid chaperonin 60 which contains two related subunits¹²). There is clearly some commonality of structure between all these particles from prokaryotes and eukaryotes. Is there also a commonality of function for protein folding?

From previous observations it seemed that TCP-1 might be involved in the assembly of microtubules, because the protein is especially abundant during spermatogenesis and mutations in this protein affect the workings of the mitotic spindle in yeast. The new results show that freshly synthesized chains of chicken α - and β -tubulin made separately, or together in a rabbit reticulocyte extract, become transiently bound to a 900K particle¹. The bound tubulin chains are

sensitive to added protease, but they can be released from the particle by ATP hydrolysis in a more protease-resistant form competent to produce α - β -tubulin heterodimers and to enter microtubules. Independent work with an 800K particle purified from reticulocyte lysate found that this particle binds β -actin chains diluted from denaturant and releases them in correctly folded form on addition of ATP³.

Similar characteristics

These characteristics are similar to those reported for a variety of other polypeptides bound to the GroEL chaperonin 60 from *Escherichia coli*⁵⁻⁷, and are interpreted in the same way — that is, that the particle functions by binding a protease-sensitive 'unfolded' form of the polypeptide, which is then released on ATP hydrolysis in a more folded, protease-resistant conformation competent for further assembly. The purpose of this train of events in the case of chaperonin 60 seems to be to prevent the 'unfolded' form of the polypeptide from aggregating to useless products. Evidence that this is also the case for tubulin is not presented, but β -actin does aggregate in the absence of the particle on dilution from denaturant³.

Analysis by SDS gel electrophoresis and immunoblotting of the particles purified from the rabbit reticulocyte lysate reveals a 58K polypeptide reactive with a monoclonal antibody against murine TCP-1; several unrelated polypeptides in the mass range 55–62K are also present in these particles, as in the case of the TCP-1 particle from murine and human cells. Whether the reticulocyte lysate particles are associated with hsp70 or chaperonin-10-like proteins is not reported; these are important points to clarify, given the evidence that chaperonin 60 binds to chaperonin 10 during its function⁷, and the proposal that hsp70 and associated proteins cooperate sequentially with the chaperonins to assist the folding of nascent polypeptides⁸. It could be that the chaperonin 60 and TCP-1 particles are part of still larger cytosolic complexes which act as protein-folding machines during protein synthesis. ►

1. Yaffe, M. B. et al. *Nature* **358**, 245–248 (1992).
2. Lewis, V. A., Hynes, G. M., Zheng, D., Saibil, H. & Willison, K. *Nature* **358**, 249–252 (1992).
3. Gao, Y., Thomas, J. O., Chow, R. L., Lee, G.-H. & Cowan, N. J. *Cell* **69**, 1043–1050 (1992).
4. Ellis, R. J. *Nature* **328**, 378–379 (1987).
5. Ellis, R. J. & van der Vies, S. M. *A. Rev. Biochem.* **60**, 321–347 (1991).
6. Gething, M.-J. & Sambrook, J. *Nature* **355**, 33–45 (1992).
7. Zellstra-Ryalls, J., Fayet, O. & Georgopoulos, C. *A. Rev. Microbiol.* **45**, 301–325 (1991).
8. Langer, T. et al. *Nature* **356**, 683–689 (1992).
9. Gupta, R. S. *Biochem. Int.* **20**, 833–841 (1990).
10. Ellis, R. J. *Science* **250**, 954–959 (1990).
11. Trent, J. T., Nimmesgern, E., Wall, J. S., Hartl, F.-U. & Horwich, A. L. *Nature* **354**, 490–493 (1992).
12. Martel, R., Cloney, L. P., Pelcher, L. E. & Hemmingsen, S. M. *Gene* **94**, 181–187 (1990).

These observations lead me to propose that the chaperonins should be redefined as a group of sequence-related molecular chaperones containing two distinct subclasses: the GroE subclass found in eubacteria, mitochondria and plastids, and the TCP-1 subclass found in archaeobacteria and the eukaryotic cytosol. Both subclasses assist the folding of newly synthesized polypeptides. How

they do this, and whether they do so in the same manner, is not clear, and the next task is to flesh out the detail on this new addition to our view of protein folding, arguably the most important single process in biology. □

John Ellis is in the Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK.

PALAEOLOGY

Wandering across time

Michael J. Novacek

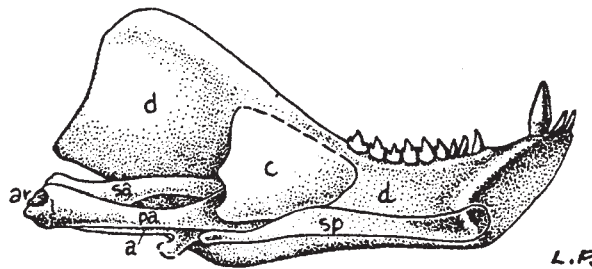
DOES a new fossil discovery, described on page 233 of this issue¹ by Fox and colleagues, extend the record of the mammal-like reptiles forward by 100 million years? The possibility is likely to attract some lively discussion.

Few fossil groups fulfil our expectations as well as the synapsids, the mammal-like reptiles. The group is richly represented by skulls and skeletons that document an orderly branching sequence leading to mammals². Especially illuminating are the therapsids, advanced synapsids occupying the neighbouring branches of the mammalian clade. Therapsids reveal the transitional states that presage the important evolutionary innovations found in mammals³. The lower jaw becomes increasingly dominated by the dentary, the sole element of the mammalian lower jaw but one of several in reptiles, and the primitive amniote post-dentary elements are reduced and more loosely articulated (see figure). The coronoid process on the lower jaw becomes more prominent and, as in mammals, is an important site for attachment of the jaw muscles. The secondary palate expands and the cheek teeth show more complexity, hinting at the mammalian capacity to carry on respiration while chewing food. In addition, the first vertebrae of the neck region show modifications that approach the definitive atlas-axis complex in mammals.

The oldest therapsids are found in early Permian faunas dating back to 280 million years ago⁴. The group is particularly well sampled from beds in western Russia, southern Africa and southern South America, and overlapped in history with mammals, whose earliest members appear some 220 million years ago⁵. Unlike the mammals, however, therapsids were not known to outlast the 160-million-year-old mid-Jurassic — at least not until the discovery of what Fox *et al.* identify as a member of the group. The

fossil was recovered from the upper Palaeocene Paskapoo Formation in Alberta, Canada, an occurrence that, if attributed correctly, extends the range of therapsids by an extraordinary 100 million years.

The newly discovered taxon, named *Chronoperates* (meaning 'time-wanderer') *paradoxus*, is represented only by part of a jaw with three posterior teeth and two empty tooth sockets or alveoli (see Fig. 1 on page 233). Yet



View of the inside face of the lower jaw of the Triassic mammal-like reptile *Cynognathus*, taken from Alfred S. Romer's classic *Vertebrate Paleontology* (University of Chicago Press, 1945). Compare this with the same view of the newly found fossil *Chronoperates* in Fig. 1b of the paper by Fox *et al.* (page 233). Abbreviations for bones: a, angular; ar, articular; c, coronoid; d, dentary; pa, prearticular; sa, surangular; sp, splenial.

palaeontologists can often retrieve essential clues from fragments, and this case is no exception. As is typical of therapsids, each tooth is single-rooted with transversely narrowed, multi-cusped crowns that lack any low shelving (cingula). In addition, the jaw architecture suggests the classic therapsid mosaic of features. Although the dentary is well developed in *Chronoperates*, this animal must have also retained the more primitive post-dentary elements. These elements are not preserved, but their presence is indicated by troughs and articular surfaces at the rear of the dentary.

The authors consider other groups to which *Chronoperates* might belong, among them the primitive symmetrodont mammals, a few advanced mammals, some lizards, and aberrant crocodilians.

In a skilfully developed argument, they eliminate these possibilities against a checklist that emphasizes microstructure of the enamel, the condition of the tooth roots and construction of the coronoid process. Nonetheless, the assignment to the therapsids is likely to attract some debate. The distinctions from symmetrodonts in particular do seem, at least to me, rather subtle. Some symmetrodonts, like therapsids, have post-dentary elements, although in symmetrodonts the teeth are double-rooted. The crowns of the teeth in *Chronoperates* compare closely with teeth in symmetrodonts⁶ as well as therapsids. Symmetrodonts are known from upper Cretaceous sediments⁵, so extension of the range of the group into the upper Palaeocene would be less dramatic.

Range extensions of fossils are frequently discovered, but this claim for late survivorship is, by any measure, a remarkable one. Therapsids are not a depauperate group, and the rich Mesozoic and Early Cenozoic faunas on several continents might be expected to produce them. So why this apparent 100-million-year hiatus? Fox *et al.* attribute the gap to a drastic post-Jurassic decrease in therapsid diversity, a geographical range that contracted to northern regions, and limited sampling with techniques that are best able to recover small, inconspicuous fossils.

It is, nonetheless, surprising that many Cretaceous faunas containing tiny mammals⁷ have failed to yield these therapsid survivors. There is also the nagging possibility that the assignment of *Chronoperates* to the therapsids is not decisively confirmed. At present, however, we can only go on the evidence available, and the evidence at least allows the authors' argument for a highly unexpected extension in temporal range for the group. At any rate, the news may send many Palaeocene *aficionados* back to the field for a second look. □

Michael J. Novacek is at the American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024, USA.

1. Fox, R. C., Youzwyshyn, G. P. & Krause, D. W. *Nature* **358**, 233–235 (1992).
2. Gauthier, J., Kluge, A. G. & Rowe, T. *Cladistics* **4**, 105–209 (1988).
3. Hopson, J. A. & Barghusen, H. R. in *The Ecology and Biology of Mammal-like Reptiles* (eds Hotton, N., MacClean, P. D., Roth, J. J. & Roth, E. C.) 83–106 (Smithsonian Institution Press, Washington, D.C., 1986).
4. Laurin, M. & Reisz, R. R. *Nature* **345**, 249–250 (1990).
5. Carroll, R. L. *Vertebrate Paleontology and Evolution* (Freeman, New York, 1988).
6. Prothero, D. R. *Bull. Am. Mus. nat. Hist.* **167**, 281–325 (1981).
7. Clemens, W. A. *Univ. Calif. Publ. geol. Sci.* **94**, 1–102 (1973).

Natural Josephson junctions

Colin Pegrum

FROM the earliest days of high-temperature superconductivity, the race has been on to develop Josephson junctions in the new materials. These are needed for many active small-scale devices, such as SQUIDS (superconducting quantum interference devices), voltage standards, microwave devices and digital logic circuits¹. There is still no known way of making true tunnel junctions, although several structures have been prepared that have non-ideal Josephson properties², some of which are good enough for commercial use³. But new experiments by Kleiner and colleagues⁴ show that Josephson tunnelling can occur *within* single crystals of superconducting $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ (BSCCO).

The Josephson effect arises where two samples of superconductor are separated by a thin layer of insulating material. The Cooper pairs that carry the supercurrent can tunnel through this barrier, so a current can flow at zero voltage, up to a critical current I_c ; this is the d.c. Josephson effect. Above I_c the flow of alternating currents (the a.c. Josephson effect) gives rise to highly characteristic current-voltage properties.

In common with all high-temperature superconductors, the samples of BSCCO have a layered structure. The planes of Cu and O (the *a-b* planes) are separated in the perpendicular direction, *c*, by about 1.2 nm, and have layers of BiO and SrO in between (these are non-superconducting, and are either metallic or insulating in nature).

Along the *c*-axis the coherence length ξ is only about 0.1 nm, which means that, for superconductivity in this direction, the wavefunction is much weakened within the intervening BiO/SrO layers. Thus the superconducting planes are only weakly coupled to each other. Another way of describing this is to view each CuO plane as a superconducting electrode, with Josephson coupling to each neighbour; this is the Lawrence–Doniach model⁵. To investigate this, Kleiner *et al.* took a small single crystal of BSCCO, as thin as possible in the *c*-axis direction, and attached contacts to the outermost *a-b* planes. The crystal, they argued, should act like a series array of Josephson junctions.

But is the coupling between layers truly Josephson-like? The authors believe it is, and set out the necessary criteria that must be met to prove it, and experiments to match. They use a tiny crystal, $30 \times 30 \mu\text{m}^2$ in the *a-b* direction, and only $3 \mu\text{m}$ in the *c* direction, which they expect to contain about 2,000

Josephson junctions. Evaporated gold contacts are made to the *a-b* faces, low enough in resistance to permit two-terminal measurements. The temperature is variable from 4.2 K upwards. A magnetic field and microwaves at a few gigahertz can be applied, and microwave emission at 2–18 gigahertz can be detected.

For true Josephson behaviour, the current-voltage (*I-V*) characteristic must have some distinguishing features; most notably, d^2I/dV^2 should be positive just above the critical current I_c . If flux motion were involved (as it is frequently for pseudo-Josephson structures), it would be negative. Kleiner *et al.* find the correct shape, though the results are at first sight confused by the existence of many different branches to the *I-V* characteristic above I_c . (These can be explained plausibly if one assumes that not all of the 2,000 junctions choose to switch to the voltage state at any one time. Perhaps more studies are needed of this effect.) But the observations go some way to confirming true Josephson tunnelling.

Another test calls for a well-defined modulation of I_c in a magnetic field *H*. The key point here is to ensure that the field is absolutely parallel to all the *a-b* planes, otherwise it can create 'Abrikosov' current vortices in those planes and cause spurious effects. Careful sample mounting ensures this. Ideally a Fraunhofer-like modulation should be seen, with $I_c(H)$ scaling as $|\sin(x)/x|$, where *x* is proportional to the flux threading its way between adjacent *a-b* planes. In practice, the authors do not see exactly this, but the size and form of the reduction of I_c either side of $H = 0$ is quite convincing. To be fair, few high-temperature junctions with apparent Josephson properties have true $|\sin(x)/x|$ behaviour.

Next comes the $I_c(T)$ test: how does I_c vary with temperature *T*? If the barrier between the superconducting (S) layers is an insulator (I), this SIS junction should have $I_c(T)$ described by the Ambegaokar–Baratoff relation. If it is a normal metal (SNS), then coupling is not by Josephson tunnelling but by the proximity effect, and a different temperature dependence applies. They see only a poor fit to Ambegaokar–Baratoff, but do calculate the temperature-dependent energy gap $\Delta(T)$ from their data and claim good agreement with measurements by others. It would be good to see further studies of $I_c(T)$ and some fits to different models, which might give a better picture of interlayer coupling.

Last, and most important, all good Josephson junctions should interact with microwaves in a unique way. A single junction irradiated at a frequency *f* will show constant-voltage steps in its *I-V* curve at $V_n = n\phi_0 f$, where ϕ_0 is the flux quantum and *n* an integer. *N* similar junctions in series can lock together in phase, and then $V_n = Nn\phi_0 f$. The authors see some evidence of this behaviour, though *N* is only 10–20, indicating that only small groups of junctions phase lock together. This is convincing proof that the sample contains a number of similar Josephson junctions.

The dependence of step height on microwave power is also much as expected. More interestingly, the authors see microwave emission from their sample when it is biased with a direct current — both broadband, when junctions are acting independently, and also peaking around several frequencies when the junctions phase lock to produce more coherent radiation. They show data for a line at 11 gigahertz. There is some evidence that the linewidth is reduced as expected by $1/N$, where *N* junctions are involved, though it is not possible to determine confidently the value of *N*. Nonetheless this is a significant demonstration of true Josephson behaviour which few other types of high-temperature junctions have shown.

The authors have therefore clearly demonstrated Josephson coupling between the CuO planes in BSCCO. Whilst their findings do not form the basis of a direct method to engineer true junctions for device applications, they give some insight into the mechanism of interlayer superconductivity in the material. It will be interesting to see these experiments repeated with other high-temperature superconductors, especially YBCO — currently the favourite for thin-film devices. The results do suggest that the way forward for making Josephson junctions must be to form layered structures in a controlled manner, by adapting established deposition techniques using, for example, molecular beam epitaxy⁶, laser ablation⁷ or sputtering⁸. Then maybe a real high-temperature Josephson tunnel junction will be made for the first time. □

Colin Pegrum is in the Department of Physics and Applied Physics, University of Strathclyde, Glasgow G4 0NG, UK.

1. Van Duzer, T. *IEEE J. Quantum Elec.* **25**, 2365–2377 (1989).
2. Rowell, J. M. *Supercond. Sci. Technol.* **4**, 692–695 (1991).
3. Char, K. *et al.* *Appl. Phys. Lett.* **59**, 733–735 (1991).
4. Kleiner, R. *et al.* *Phys. Rev. Lett.* **68**, 2394–2397 (1992).
5. Appel, J. & Fay, D. *Phys. Rev.* **B41**, 873–876 (1990).
6. Virshup, G. F. *et al.* *Appl. Phys. Lett.* **60**, 2288–2290 (1992).
7. Li, Q. *et al.* *Phys. Rev. Lett.* **64**, 3086–3089 (1990).
8. Triscone, J.-M. *et al.* *Phys. Rev. Lett.* **64**, 804–807 (1990).

Still poles apart on reversals?

Andy Jackson

TREMENDOUS interest has been shown recently in determining how the Earth's magnetic field changes polarity. On pages 226 and 230 of this issue, Langereis *et al.*¹ and Constable² probe two aspects of the most intriguing suggestion about the reversal: that the Earth's dipole, or virtual geomagnetic pole (VGP), flips in a particular way by following preferred meridional paths during a reversal.

Langereis *et al.*, considering the way in which sedimentary sequences record geomagnetic reversals, present a straightforward explanation for this simple flipping picture, which may be an artefact, they propose, of the pre- and post-reversal positions of the VGP. And Constable upturns all conventional thinking by showing that, over the past 5 million years, there has been a discernible tendency for the stable field (normal or reversed) VGPs to be biased towards two longitudinal bands rather than being uniformly distributed in longitude around the geographical North Pole, and that these are the very same bands that frequently occur in the reversal records. Intriguingly, this bias points to control of the magnetic field (generated in the Earth's core) by the mantle. These two observations are complementary and somewhat unifying: if many sedimentary reversal records do depend on pre- and post-reversal VGP positions, then preferential stable-field VGP positions would lead to a statistical predominance of reversal paths apparently along the same longitude bands, at least if reversals happen randomly.

It is well known that the present-day magnetic field at the Earth's surface is well approximated by a dipole tilted with respect to the rotation axis, and that the

contribution from higher multipoles is small. If the field were entirely dipolar, then observers taking measurements of declination (the angle between a compass needle and true north) and inclination (the angle between a needle and horizontal) all over the globe would each calculate the same dipole as their source. The position that the south pole of this dipole cuts the Earth's surface is the virtual geomagnetic pole (Fig. 1). A consequence of higher multipole components in the field would be to cause a scatter in the VGPs which each observer computes. Normal-polarity states have VGPs in the Northern Hemisphere (present day VGPs are clustered over Greenland and Canada) and reversed-polarity states have VGPs in the Southern Hemisphere. That the field has flipped states randomly on a million-year time-scale is shown for the past 200 million years (Myr) by the magnetic stripes on the seafloor, which grows as new hot rock is extruded at mid-ocean ridges and freezes in the instantaneous polarity of the field as it cools.

It is not so easy to document a transition (perhaps lasting just 10,000 years) between stable polarities. Although they are generally accurate recording media, lava flows on land tend to be sporadic, so that capturing a reversal is rather difficult, and precise dating is problematical. The simplest way is to examine the magnetization locked into continuously deposited sedimentary sequences; the youngest and best documented reversal, the Matuyama-Brunhes (730,000 yr ago), is found in many recent sediments. The excitement generated by these reversal records is a result of the mounting evidence³⁻⁵ that the VGPs calculated for states intermediate between the normal and reversed states (transitional states) tend to be preferentially confined to two antipodal bands of longitude, one centred over the Americas and the other centred over eastern Asia and Australia. Certainly the evidence is visually appealing⁴, and controversial⁶. One of the main points of contention is that different data recording the same reversal can give different VGP paths, even at the same site, essentially nullifying the hypothesis that the field remains dipolar during the transition. It is these types of data with which Langereis *et al.* concern themselves.

Certainly sediments are an imperfect recording medium for reversals. The processes by which sediments acquire their magnetization are not well understood, and are thought to depend on a variety of factors including magnetic

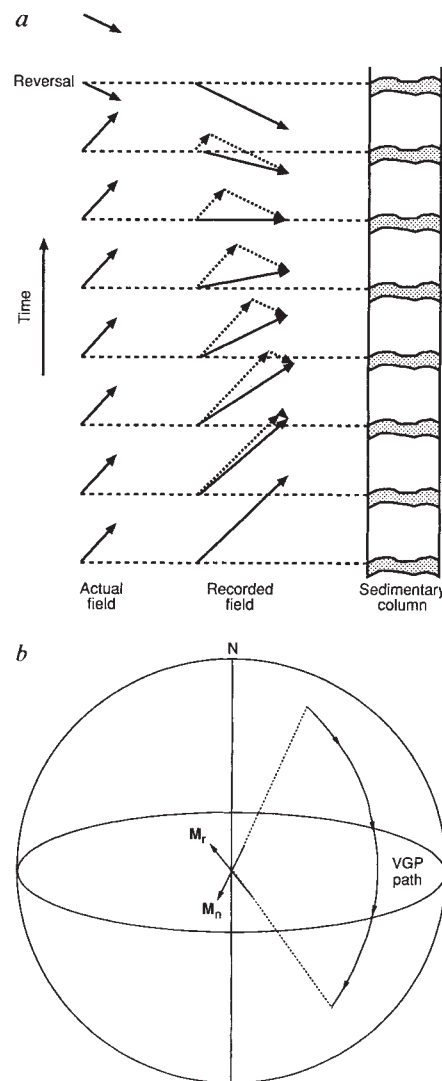


FIG. 2 Langereis *et al.*'s model for recording of a reversal in their sedimentary sequence. *a*, Schematic diagram for two field states taken to be coplanar for simplicity. Because a rock continues to record the magnetic field for a substantial time after it is laid down, the magnetization recorded depends on the field both at the time of deposition and at times afterwards (stippled arrows); the last components added are the ones on the right. The net magnetization we measure at a particular point in the sedimentary column (solid arrow) is the vector sum of the two field components. *b*, The VGP path which would be attributed to the reversal from dipole M_n to M_r from the sedimentary record.

mineralogy, duration of the transitional field state and sediment deposition rate. One thing is known: sediments often smooth the transition on a timescale generally long with respect to the transition, as demonstrated by the fact that changes in intensity are recorded over a longer timespan than changes in direction⁷.

Langereis *et al.* examine marls from the Mediterranean which document several reversals over the past 7 Myr. They measure pairs of (normal, reversed) VGPs immediately before and

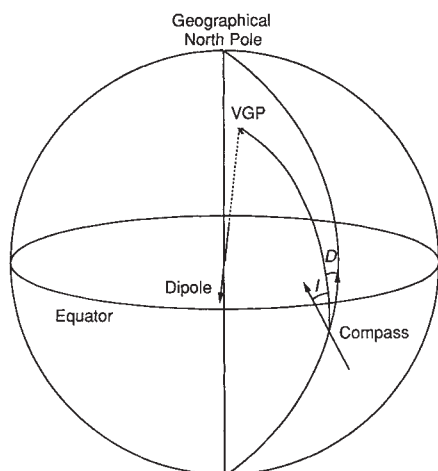


FIG. 1 Declination (D) and inclination (I) measurements predict a VGP, which removes the dependence on the position of the site.

after the reversal, poles which they call near-transitional (see also ref. 8). They test whether these pairs are antipodal, finding strong evidence that they are not. This is not too surprising, as only rather special reversals would lead to antipodality. They use these poles to simulate the VGP reversal paths which would be recorded when one VGP pole is gradually replaced over time by the VGP pole of the opposite polarity, attempting to mimic the acquisition of sedimentary remanence over an interval that is long compared with the reversal duration (Fig. 2).

Actually, the analysis is more subtle than Fig. 2 suggests. In addition to measuring near-transitional pairs of VGPs, the authors measure VGPs and simulate transitions from stable field states well away from the reversal. The results are rather impressive, at least for their particular rocks: nearly all the reversal data correlate well with the synthetic data formed by averaging the near-transitional poles, although occasionally they appear to follow the stable-field versions. The implication is, then, that sedimentary records tell us more about the pre- and post-transitional VGPs than about the actual reversal — it also helps to explain how we can observe different (for example, antipodal) VGP paths for a single reversal, because, depending on the sedimentary recording mechanism, we can record the VGP path from either of two pairs (stable or near-transitional) of poles.

Constable's analysis concentrates on the record of stable polarity direction measurements recorded in lavas over the past 5 Myr, using the dataset compiled by Lee⁹. Some of these data have been available since the early 1960s, and they were shown to give VGPs which were both 'far sided' (beyond the geographical pole with respect to the observation site) and 'right handed' (having positive declinations)^{10,11}. However, previous analyses generally concluded that the data gave VGPs which were equally scattered in longitude around the geographical North Pole. The remarkable picture from Constable's analysis is that, contrary to popular belief, the individual VGPs (over 2,000 of them) are actually biased towards two longitudes, and these are the same longitudes seen predominantly in the reversal records. This conclusion seems to be insensitive to both the distribution of sampling sites and also to the effects of recent plate motions. Constable also notes that the present-day VGPs cluster in one of these preferred bands, and a reversal now would track through the Americas.

Where does this leave us? Both analyses lead to the conclusion that the magnetic field has preferred orientations in space. The obvious candidate for NATURE • VOL 358 • 16 JULY 1992

directional control is the mantle, because convection in the mantle occurs on a timescale that is long compared to that for convection in the core. Given perfect spherical geometry, the convective motions in the core responsible for field generation would tend to drift around the rotation axis, and it has been conjectured that anomalies in temperature or heat flux at the core-mantle boundary could prevent this drift¹². Seismic velocity anomalies at the base of the lower mantle certainly provide evidence for inhomogeneities, which could be interpreted in terms of temperature differences. This conjecture has now been put on a firm mathematical footing, for convection without magnetic fields at least. Zhang and Gubbins¹³ find that drifting convection rolls can indeed become locked into anomalies in temperature on the boundaries for long periods of time, provided the phase of convection and of

the boundary anomalies is correct. Clearly a firming-up of recent ideas, as well as some old ones, is happening. □

Andy Jackson is in the Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK.

1. Langereis, C. G., van Hoof, A. A. M. & Rochette, P. *Nature* **358**, 226–230 (1992).
2. Constable, C. *Nature* **358**, 230–233 (1992).
3. Tric, E. *et al. Phys. Earth planet. Inter.* **65**, 319–336 (1991).
4. Laj, C. *et al. Nature* **351**, 447 (1991).
5. Clement, B. M. *Earth planet. Sci. Lett.* **104**, 48–58 (1991).
6. Valet, J.-P., Turcholka, P., Courtillot, V. & Meynadier, L. *Nature* **356**, 400–407 (1992).
7. Hillhouse, J. & Cox, A. *Earth planet. Sci. Lett.* **29**, 51–64 (1976).
8. Hoffman, K. A. *Nature* **354**, 273–277 (1991).
9. Lee, S. thesis, Australian National Univ. (1983).
10. Wilson, R. L. *Geophys. J. R. astr. Soc.* **19**, 417–437 (1970).
11. Merrill, R. T. & McElhinny, M. W. *The Earth's Magnetic Field* (Academic, London, 1983).
12. Bloxham, J. & Gubbins, D. *Nature* **317**, 777–781 (1985).
13. Zhang, K. & Gubbins, D. *J. Fluid Mech.* (submitted).

FULLERENES

Down the straight and narrow

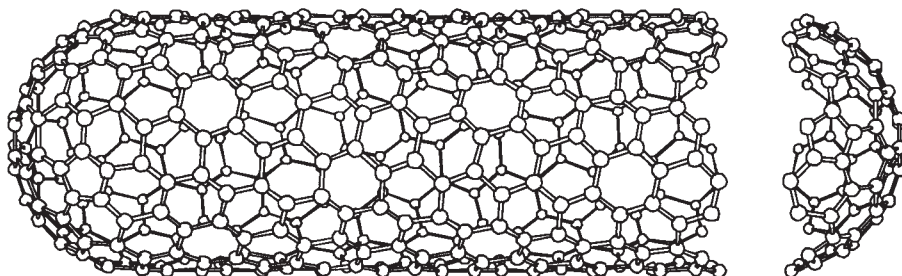
M. S. Dresselhaus

JUST as the discovery by Krätschmer *et al.*¹ of a method to synthesize fullerenes in gram quantities heralded a flood of activity in the study of C₆₀, the report of the large-scale synthesis of carbon nanotubes by Ebbesen and Ajayan, on page 220 of this issue², should stimulate a flurry of experimental work on these tubular relatives of the fullerenes. In particular, the availability of macroscopic quantities of carbon nanotubes should make it possible for experimentalists to test the exotic electronic and mechanical properties recently predicted for them.

In their report, Ebbesen and Ajayan describe the production of copious quantities of carbon nanotubes embedded in a geode-like cylindrical shell. Their method appears in many ways similar to that used by Bacon in 1960 to grow graphite whiskers³. Both used a d.c. discharge between carbon electrodes (75–80 volts and 70–76 amps for Bacon compared with 18 volts and 100 amps for

Ebbesen and Ajayan), the positive electrode having a smaller diameter than the negative electrode, and both carried out the discharge in an inert gas (at the high pressure of 92 atmospheres for Bacon, whereas the new work uses a modest pressure of 2/3 atmosphere). When Bacon cracked open his hard cylindrical boules, he found, protruding from the fracture surfaces, scroll-like carbon whiskers up to 3 cm long and 1–5 µm in diameter. These revealed great crystalline perfection, and had high electrical conductivity and elastic modulus. Since their discovery, graphite whiskers have been the standard by which the performance of carbon fibres is measured.

In the area of nanotubes, the present standards for comparison have been theoretical models. Calculations have been done, in our laboratory and in many others, on graphene nanotubes^{4–6}, consisting of a planar honeycomb network of carbon atoms on a graphite sheet one atom thick and rolled up into a



Theoretical model of a single-layered carbon nanotube showing caps at each end and the tube's chirality.

cylinder. These nanotubes have remarkable properties when the diameter of the cylinder is comparable to that of the C_{60} molecule and the length is much greater. Each nanotube is specified by its diameter and the chirality of the atomic rows of carbon atoms relative to the cylinder axis (see figure showing atomic arrangement and chirality for a graphene nanotube). The 2π periodicity of the cylinder restricts the number of allowed electron wave vectors in the circumferential direction, the allowed number depending on the fibre diameter and chirality.

Calculations show that approximately a third of the possible tubes with diameters comparable to those of typical fullerenes are metallic⁴⁻⁶, as is also the case at room temperature for a planar graphite sheet one atomic layer in thickness. However, for the remaining two-thirds of the possible thin nanotubes, an electronic band gap forms. Thus theory predicts that whether a graphene nanotube is metallic or not is determined by geometric considerations.

Such a situation is exceptional in solid-state physics. The availability of carbon nanotubes should allow the prediction to be tested by, for example, scanning-tunnelling-microscope (STM) spectroscopy for a single-layer tube, whose diameter and chirality might also be measured with the same microscope. More detailed tests of these theories could be provided by varying the voltage of the STM tip relative to that of the tube, thereby providing information on the density of states near the electrons' characteristic Fermi energy for various fibre geometries.

Strength

Calculations of the mechanical properties⁷ suggest that carbon nanotubes are significantly stiffer than currently available carbon fibres or, in fact, than any presently known material. Furthermore, the strong sp^2 bonding and very small C-C distance would appear to inhibit the introduction of impurities and defects into the nanotubes, thereby offering the potential of great tensile strength. Experimental confirmation of these predictions would be significant in establishing the ultimate performance limits for carbon fibres⁸, a technologically important commercial material in the space age.

While calculations have focused on single-layer graphene tubes with diameters comparable to that of C_{60} , experimental observations^{2,9,10} have mainly involved specimens comprising two or more concentric tubes, and with diameters of more than 20 Å. Because the semiconducting gap between the electronic valence and conduction bands decreases approximately as the inverse of

the fibre diameter, experimental work will probably be mostly on tubes with the smallest available diameters, where the geometric effects are predicted to be largest and the most interesting behaviour is expected. Calculations done in our group on bilayered tubes provide some justification for relating the concentric tubes that are now experimentally available to the electronic properties predicted for single-layer tubes: interactions between concentric adjacent tubes do not apparently destroy the metallic or semiconducting properties of the constituent layers.

Growth

The availability of copious amounts of carbon nanotubes will not only help in the exploration of their exotic electronic structure and mechanical properties, but should also stimulate study of the way they grow. In essence, the growth of fullerene balls and tubes involves the addition of a sequence of individual carbon hexagons. If, as claimed (M. Endo and H. Kroto, personal communication), the accretion of carbon atoms onto a tube is catalysed at a pentagonal face in the tube's end cap (see figure), then the growth mechanism should also be related to the growth mechanism for general fullerenes. STM studies of carbon nanotubes in the cap region or near an open end (if such exists) may be a good way to elucidate fullerene formation and growth.

Lastly, it is intriguing to think that nanoscopic electronic devices could be worked up from concentric semiconducting and metallic carbon tubes. These could function as capacitors in memory devices or as transistors in switching circuitry. Because the tubes do not require any doping by impurities, as conventional semiconductors do, but gain their electronic properties from their geometry, the resulting devices should be highly thermally stable and have high intrinsic mobility. □

M. S. Dresselhaus is in the Department of Electrical Engineering and Computer Science, and the Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

1. Krätschmer, W., Lamb, L. D., Fostiropoulos, K. & Huffman, D. R. *Nature* **347**, 354–358 (1990).
2. Ebbesen, T. W. & Ajayan, P. M. *Nature* **358**, 220–222 (1992).
3. Bacon, R. J. *Appl. Phys.* **31**, 283–290 (1960).
4. Saito, R., Fujita, M., Dresselhaus, G. & Dresselhaus, M. S. *Appl. Phys. Lett.* **60**, 2204–2206 (1992).
5. Mintmire, J. W., Dunlap, B. I. & White, C. T. *Phys. Rev. Lett.* **68**, 631–634 (1992).
6. Hamada, N., Sawada, S.-I. & Oshiyama, A. *Phys. Rev. Lett.* **68**, 1579–1582 (1992).
7. Overney, G., Zhong, W. & Tománek, D. (preprint).
8. Dresselhaus, M. S., Dresselhaus, G., Sugihara, K., Spain, I. L. & Goldberg, H. A. in *Graphite Fibers and Filaments* (Springer, Berlin, 1988).
9. Iijima, S. *Nature* **354**, 56–58 (1991).
10. Endo, M., Fujiwara, H. & Fukunaga, E. in *Proc. Second C₆₀ Symp., Japan*, 101–104 (Japan Chem. Soc. 1992).

Space mothballs

Big orbiting mirrors to focus sunlight down to the ground have often been suggested. Daedalus now proposes a novel orbiting lens.

His first idea was a transparent space balloon, inflated to low pressure with some refractive gas. A spherical balloon lens would be easier to make than a conventional lenticular one, and would not have to be oriented to face the Sun; but spherical lenses have dreadful optical aberrations. The Luneberg spherical lens gets round this problem. It is cunningly made of concentric layers, whose refractive indices drop off from the centre to the outside.

The obvious way to implement a Luneberg gas lens in space is to introduce gas steadily in the middle, and let it flow and expand in all directions out into the surrounding vacuum. The refractive index will then drop off radially, with the density of the outflowing gas. Even better, the outer envelope is then unnecessary. The lens becomes an enormous tenuous gas cloud spreading continuously from a tiny evaporating nucleus — which need only be a block of some volatile solid like naphthalene.

So Daedalus's proposed space lens is simply a large naphthalene mothball. Once in orbit, it will develop a huge Luneberg-lens atmosphere of expanding naphthalene vapour, hundreds of kilometres across, intercepting gigawatts of sunlight but deviating it very weakly. The lens will have a focal length of hundreds or thousands of kilometres, comparable with its orbital height. It will focus a solar hotspot on the Earth beneath.

No feasible orbit could hold this hotspot geostationary. Daedalus plans an oblique orbit to write a 'raster' on the rotating Earth beneath: not a narrow trail of fire and destruction, but a wider, less focused band. Each pass of the mothball would bring a fiercely bright five minute heatwave, with equivalent brief twilit cold snaps on either side from where the sunlight has been deviated.

A set of orbiting mothballs, each replaced when it has nearly evaporated, will bring these sunlight pulses regularly to most points on Earth. Plant life everywhere will benefit; for photosynthesis is more efficient in interrupted light. Global weather will also improve. At each pass of each mothball, the cold phases will trigger thick clouds into raining, while the hot phase will burn off thin cloud. Countries which now only have climate will enjoy real weather! Solar energy could be locally captured too. Thermal collectors, timed to open briefly as each mothball passed overhead, could accumulate free heat for hot water, and free cold for air conditioning.

David Jones

Bear conservation genetics

SIR — We wish to report that hairs collected in the field are a potentially valuable source of DNA for genetic studies that may help in the management of the endangered Pyrenean brown bear (*Ursus arctos*) without disturbing the animals themselves.

Recent estimates suggest that the Pyrenean bear population consists of only 9–13 individuals. It may be possible to avoid extinction (the permanent loss of the Pyrenean genotypes) by reinforcing this endangered population with brown bears of other European origin, provided that the genotypes are not too different. For this reason, the French Ministry of the Environment has started a genetic study on the phylogenetic relationships among the different European populations, using the analysis of mitochondrial (mt) DNA sequences^{1,2}. The most difficult problem is the sampling of the Pyrenean population itself: the capture of wild bears by anaesthesia to obtain tissue for genetic analysis entails disturbance and risks, while their mountainous habitat is a severe limitation.

Higuchi *et al.*³ used human hair roots as a source of DNA for genetic analysis via the polymerase chain reaction (PCR). We sought to use this same source of DNA for bears, with hairs collected in the field. Indeed, it is rela-

tively easy to obtain hairs from a small piece of wire netting nailed to a tree trunk where a bear occasionally scratches.

DNA was extracted from the roots of single hairs¹ and approximately 2.5% was subjected to PCR to amplify a part of the mtDNA control region. After gel purification, an asymmetric PCR was performed to generate single-stranded DNA suitable for direct sequencing (see figure). Comparison of 269 base pairs of the mtDNA control region gives a genetic distance ($d \pm \text{s.d.}$) of $7.01 \pm 1.67\%$ between a Pyrenean and a Romanian bear, showing that this region is variable enough to reveal intraspecific variation.

Further, analysis of nuclear microsatellite DNA sequences is possible even using degraded DNA⁵, suggesting that in the Pyrenean brown bear, family relationships may be revealed as well as problems of inbreeding; it would also be possible to demonstrate the successful reproduction of introduced animals.

Pierre Taberlet

Jean Bouvet

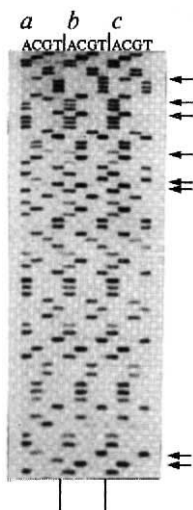
Génétique & Ecologie des Populations, Université Joseph Fourier, BP 53X, F-38041 Grenoble Cedex, France

- Wayne, R. K. & Jenks, S. M. *Nature* **351**, 565–568 (1991).
- Bowen, B. W., Meylan, A. B. & Avise, J. C. *Nature* **352**, 709–711 (1991).
- Higuchi, R. *et al.* *Nature* **332**, 543–546 (1988).
- Walsh, P. S., Metzger, D. A. & Higuchi, R. *BioTechniques* **10**, 506–513 (1991).
- Hagelberg, E., Gray I. C. & Jeffreys, A. J. *Nature* **352**, 427–429 (1991).

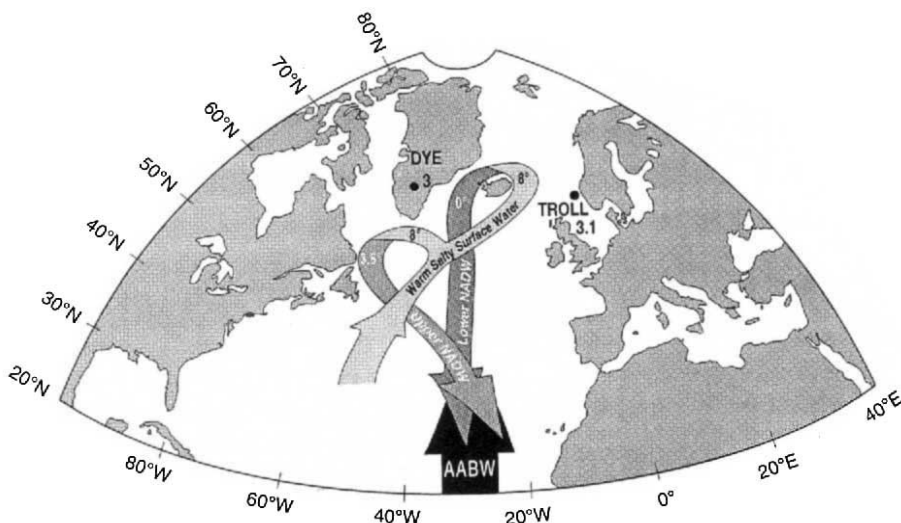
Deep circulation revisited

SIR — Although we thank Zahn¹ for his thoughtful News and Views article about the apparently conflicting evidence for changes in deep circulation during the deglaciation presented by Veum *et al.*² and ourselves³, we note that the opening sentences of his article misrepresent one of our main conclusions. Zahn states that "... Lehman and Keigwin and Veum *et al.* have examined tracers of ocean circulation in the northern corner of the North Atlantic and find that changes there ... were not responsible for sudden, temporary regressions to glacial climes". We think that we presented a strong case that they were. We did, however, propose a simple twist of the conveyor-belt/salt-oscillator hypo-

thesis⁴ which recognizes the greater climate-forcing potential of lower North Atlantic deep water (NADW) and which takes into account recent observations from the Southern Ocean that NADW reached that area during the cold Younger Dryas period (~11,000–10,000 ¹⁴C yr BP)⁵. As summarized by Zahn, we argued that the present day conveyor has, in effect, two subsurface belts which carry NADW southwards (see figure). The upper belt is fed primarily by open ocean convection in the Labrador Sea⁶ and the lower one primarily by convection within the Norwegian Sea basin⁷. Although sea surface temperatures at the two source areas are nearly identical at the time of convection, waters feeding



Autoradiography of a sequencing gel showing a variable part of the mtDNA control region for three brown bears. The sequences were obtained after amplification and direct sequencing using single hairs as a source of DNA. *a*, Killed individual from Romania (hair provided by L. Kalabér); *b*, wild individual from Pyrenees (hair collected in the field by J.-J. Camarra and G. Caussimont; the size of the tracks allowed a precise identification of an old male named "Papillon"); *c*, captive individual from Abruzzo, Italy (hair provided by G. Boscagli). Arrows, variable sites.



Schematic representation of the components of NADW which can be traced directly to source areas at the surface of the northern Atlantic. Additional components, which together amount to approximately half of the total southward flux of NADW, result from entrainment of recirculated waters, and are not shown. Temperatures (in °C) refer to those at the surface before convection, and to the average for each of the convecting water masses which feed the belts. At times in the past when the lower belt ceased operating, a greater portion of the deep Atlantic was influenced by Antarctic Bottom Water (AABW).

the upper belt cool only half as much before sinking. Thus, feeding of the lower belt releases twice as much heat to the atmosphere per unit mass, making it a more effective climate-forcing agent.

It is well documented that the lower belt ceased operating during the height of the last glaciation (and it was consequently replaced in its depth domain by Antarctic bottom water) while the upper belt persisted in some form but shoaled slightly from its present depth^{8,9}. Under these conditions, the Southern Ocean was not strongly influenced by NADW⁵. We suggested a model in which the upper belt gradually thickened and/or deepened in response to improving surface conditions as the deglaciation progressed, explaining the overall palaeochemical trends observed in the deep Atlantic^{8,10}. Meanwhile, the lower belt came and went on brief timescales, in response to changing conditions at its source area in the Norwegian Sea. By the time of the Younger Dryas, we argued, the upper belt had deepened and/or thickened enough to influence the Southern Ocean, thus explaining the recent observations there, but the tempor-

ary absence of the lower belt led to extreme cooling at the surface in the North Atlantic region. Thus we claim that the history of the lower belt was indeed responsible for the "sudden regressions to glacial climes" which were so well recorded by our surface record (Troll 3.1) and on Greenland.

Scott J. Lehman

Lloyd D. Kelgwin

*Woods Hole Oceanographic Institution,
Woods Hole, Massachusetts 02543, USA*

1. Zahn, R. *Nature* **356**, 744–746 (1992).
2. Veum, T., Jansen, E., Arnold, M., Beyer, I. & Duplessy, J. C. *Nature* **356**, 783–785 (1992).
3. Lehman, S. J. & Kelgwin, L. D. *Nature* **356**, 757–762 (1992).
4. Broecker, W. S. *et al. Paleoceanography* **5**, 469–477 (1990).
5. Charles, C. D. & Fairbanks, R. G. *Nature* **355**, 416–419 (1992).
6. Clarke, R. A. & Gascard, J. C. *J. phys. Oceanogr.* **13**, 1764–1778 (1983).
7. Aagard, K., Swift, J. H. & Carmack, E. C. *J. geophys. Res.* **90**, 4833–4846 (1985).
8. Boyle, E. A. & Kelgwin, L. D. *Nature* **330**, 35–40 (1987).
9. Oppo, D. & Fairbanks, R. *Earth planet. Sci. Lett.* **86**, 1–15 (1987).
10. Kelgwin, L. D., Jones, G. A., Lehman, S. J. & Boyle, E. A. *J. geophys. Res.* **96**, 15,811–16,826 (1991).

■ The sense of the sentence from Dr Zahn's News and Views article cited above was inadvertently changed by *Nature's* editors.

Diamond source

SIR — Carlisle¹ provides new data concerning the carbon isotope composition of diamonds found at the Cretaceous/Tertiary (K/T) boundary^{2,3}. The $\delta^{13}\text{C}$ of these diamonds turns out to be -48‰ , which is thought to be within the range of CM2 meteorites¹. But the range quoted by Carlisle¹ and ascribed to Huss⁴ is not in the cited reference, nor is it correct.

It can be seen from the detailed study of diamonds in primitive chondritic meteorites⁵ that the $\delta^{13}\text{C}$ of these components varies from -32 to -38‰ , making the $\delta^{13}\text{C}$ value of -48‰ measured by Carlisle a very important discovery. The result is even more remarkable in light of work⁶ showing that microdiamonds from the K/T boundary at Brownie Butte, Montana, and Berwind Canyon, Colorado, have $\delta^{13}\text{C}$ values of $-18 \pm 2\text{‰}$. The carbon isotope data alone are clearly consistent with a terrestrial origin for the microdiamonds formed at these particular sites.

However, there is even stronger evidence to refute an association of diamonds from the K/T boundary with those in CM2 meteorites, namely the complete disagreement of the nitrogen isotope composition. Nitrogen from meteoritic diamonds shows a characteristic $\delta^{15}\text{N}$ value of $-343 \pm 16\text{‰}$ (ref. 5). In contrast, that contained in the K/T diamonds has a $\delta^{15}\text{N}$ of $+5.5\text{‰}$ (ref. 6).

Thus, there seems to be a discrepancy in the interpretation of diamonds from the K/T boundary. To assess the credentials of the diamonds isolated by Carlisle^{1,3}, an estimate of the likely error associated with the measured $\delta^{13}\text{C}$ value of -48‰ is needed, as well as a measure of the nitrogen isotope composition of these diamonds.

Ian P. Wright

*Planetary Sciences Unit,
Department of Earth Sciences,
Open University,
Milton Keynes MK7 6AA, UK*

1. Carlisle, D. B. *Nature* **357**, 119–120 (1992).
2. Flexer, A., Rosenfield, A., Honigstein, A. & Dvorachek, M. *Lunar planet. Sci.* **XX**, 337 (1988).
3. Carlisle, D. B. & Braman, D. R. *Nature* **352**, 708–709 (1991).
4. Huss, G. *Nature* **347**, 159–162 (1990).
5. Russell, S. S., Arden, J. W. & Pillinger, C. T. *Science* **254**, 1188–1191 (1991).
6. Gilmour, I., Russell, S. S., Pillinger, C. T., Lee, M. & Arden, J. W. *Lunar planet. Sci.* **XXIII**, 413–414 (1992).

Signal sequence identified

SIR — Wei and Cresswell¹ reported that in the mutant cell line T2, which is defective in presumptive peptide transporters and therefore in MHC class I-mediated antigen presentation, HLA-A2 molecules contain large amounts of a limited set of endogenous peptides. Two of the three peptides they sequenced are derived from the signal sequence of IP-30, a soluble protein found in both extracellular and lysosomal locations², but the origin of the third was unknown. We have now identified this sequence as a carboxy-terminal portion of the signal sequence of the human α -subunit of the so-called signal sequence receptor (SSR α):

human SSR α : MRLPLRLLLLVFPATVLRGGPRGSLAVA⁻¹ ODLTE...
peptide: VLRGGPRGLLAV⁺¹

The deviation at the -5 position could be caused by either a mutation or perhaps a sequencing error. The cleavage site of the signal sequence is tentative. The sequence of the human SSR α has not yet been entered into a database and differs in three positions in the 13-mer from the canine sequence³, so it would not have been picked up by Wei and Cresswell.

SSR α is a constituent of an abundant heterotetrameric membrane protein complex in the rough endoplasmic reticulum⁴. It is probably located in the protein-translocation site but, although it is not involved in signal sequence recognition, its precise function is unknown.

Including the results of Henderson *et al.*⁵, who found portions of the signal sequences of IP-30 and of calreticulin (a luminal endoplasmic reticulum protein) to be bound by HLA-A2, it seems reasonable to conclude that in the mutant cell line only peptides of signal sequences are bound by the MHC class I molecules.

The most plausible explanation is that in the absence of peptide transporters the only peptides present in the lumen of the endoplasmic reticulum near newly made MHC molecules are derived from cleaved signal sequences. 'Endoplasmic reticulum-degradation' (ref. 6) of other portions of proteins seems to occur elsewhere.

Enno Hartmann

Tom A. Rapoport

*Max-Delbrück-Center for Molecular
Medicine,*

*Robert-Rössle-Str. 10,
O-1115 Berlin-Buch, Germany*

Siegfried Pohn

*Institute of Biochemistry,
Humboldt-University,
Hessische Str. 3-4,
O-1040 Berlin, Germany*

1. Wei, M. L. & Cresswell, P. *Nature* **356**, 443–446 (1992).
2. Luster, A. D., Weinshank, R. L., Feinman, R. & Ravetch, J. V. *J. biol. Chem.* **263**, 12036–12043 (1988).
3. Pohn, S. *et al. Eur. J. Biochem.* **188**, 439–445 (1990).
4. Rapoport, T. A. *FASEB J.* **5**, 2792–2798 (1991).
5. Henderson, R. A. *et al. Science* **255**, 1264–1266 (1992).
6. Klausner, R. D. & Sitia, R. *Cell* **62**, 611–614 (1990).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references.

Bred in the bones

Michael Neve

Articulating the Elephant Man: Joseph Merrick and His Interpreters. By Peter W. Graham and Fritz H. Oehlschlaeger. Johns Hopkins University Press: 1992. Pp. 212. \$24.50, £18.

MODERN literary criticism, not least when applied to historical questions, has taught us an important lesson: historians are as likely to be writing a certain kind of fiction as novelists. The history of huge events such as the French Revolution, individual figures such as Elizabeth I or institutions such as the Royal Society is just as much the history of historians as it is some 'real' history that students have objectively uncovered. The most striking recent attempt to unite the historical and the literary is the work of the Harvard scholar Simon Schama. In his history of the French Revolution, *Citizens*, and in later imaginative writings, Schama has asked historians to blow their own cover and to admit that they interpret and even create events out of the factual record and that this is no fault, no act of fictional dishonesty. Of course, there are facts that do have to be incorporated. But bringing history alive partly involves admitting that acts of interpretation are necessary and that the dream of objective truth is itself a fiction in disguise.

It is these kinds of concern that inform this latest study of the remarkable case of Joseph Merrick, the so-called 'Elephant Man' of Victorian England. As the authors strikingly express in their introduction, the story of the misshapen, inarticulate and apparently unexpressive Merrick generates a kind of unfinished story, a story, furthermore, that has the power to create its interpreters. Whether a nineteenth-century surgeon, a twentieth-century anthropologist or even a twentieth-century pop star, interpreters of the 'real' Merrick all bring news about themselves, about their culture and their place in it. Indeed, one of the merits of this fine study is the implication that Merrick could be said to have come to interpret himself, knowing that he was both a subject in himself and an object to an often scared but also voyeuristic outside world. By examining the history of the interpretation of the Elephant Man, two professors of English, in the middle of a prolonged historical

freak show, restore Merrick to himself.

The best parts of the book are its earliest sections on the surgeon Sir Frederick Treves (1853–1923) and his time at the London Hospital in Whitechapel. Treves's interest in Merrick and his housing him at the London Hospital are fairly well known. The

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Joseph Merrick — the so-called 'Elephant Man' of Victorian England

authors illuminate a familiar story by reminding us of, and then examining, Treves's life as a writer. Treves not only published clinical reports on Merrick's condition, but late in life also produced a reminiscence of him, along with other stories of "queer unknown patients". With great skill, the authors show that a mixture of literary convention and medical matters had made possible the relationship between Treves and Merrick, with Treves imitating his poet hero Robert Browning in his interpretation of his patient. Treves romanticizes Merrick, stressing the gap between beastly flesh and aspiring spirit, and leaves a writer's

story for later interpretation. At Treves's own funeral, in Dorchester, Dorset, it was entirely appropriate that among those at the graveside was his fellow Dorset author, Thomas Hardy.

Later interpretations of Merrick, by anthropologists such as Ashley Montagu, or film makers such as David Lynch, share some common themes while also taking Merrick from books to the stage (which he would undoubtedly have liked) and to the big screen. All the interpretations of his life differ, especially among the playwrights, where Victorian ideas of charity and kindness receive a variety of harsher rewritings, not least by emphasizing the element of power in the doctor-patient relationship. Nonetheless, a common theme of Merrick and his feelings for women, not least a search for a lost mother and an admiration for female actresses, keeps the general story together. In ways that are familiar to students of nineteenth-century psychiatry, one might call this general story Shakespearean, taking in love and its various fortunes, as well as a powerful sense of homelessness and absent friends.

This journey, made on behalf of an invented Merrick, from the warm inside of an East London hospital to the cold outside of the twentieth-century *fin de siècle*, allows the authors to conclude with the story about the pop star Michael Jackson apparently wishing to imitate aspects of Merrick's story while also wanting to buy his bones from what is now the Royal London Hospital. This story seems rather too much that overplayed modern drama in which the billionaire tries to tell us about feeling an outsider; but the point of this admirable account is that this is the latest story and one that is telling us something about the history of stories and our place in them.

The book reminds us not least about one of the words in its title. To 'articulate' is to speak distinctively and intelligently and also to join together, even to join bones together. By examining how people have written and spoken of Merrick, by interpreting his interpreters, Graham and Oehlschlaeger have helped his bones to rest in peace. □

Michael Neve is in the Department of Anatomy and Developmental Biology, University College, Gower Street, London WC1E 6BT and at the Wellcome Institute for History of Medicine, 183 Euston Road, London NW1 2BN, UK.

The Royal London Hospital Archives and Museum

Plugging the gene pool

Norman Myers

Conservation of Biodiversity for Sustainable Development. Edited by O. T. Sandlund, K. Hinder and A. H. D. Brown. Oxford University Press/Scandinavia University Press: 1992. Pp. 327. £35.

WE hear much about mass extinction, but less about the other dimension of the biodiversity crisis, the progressive depletion of genetic variability. Suppose we witness the demise of one-third to one-half of all species within the foreseeable future, as is all too possible. The survi-

Populations for Conservation (Cambridge University Press, 1987)).

Now we have a book that specifically seeks to address this neglected area. The best section is that tackling the ecological context of genetic erosion. Papers by leading experts in the field, such as Robert Berry of University College London, Paul Harvey and his colleagues at the University of Oxford and Nils Stenseth of the University of Oslo, are outstanding in their efforts to expand our understanding of biological conservation from the standpoint of ecological genetics. They tackle key factors of genetic variability, bottleneck constraints, island biogeography, critical minimum sizes of gene stocks, population abundance in relation to population persistence and modelling predictions of ecological change. The book is worth purchasing for these papers alone.

RGS/Chris Caldwell

Other notable contributions include an opening overview of biodiversity challenges by Jeff McNeely, who presents an illuminating research agenda with specific reference to management. Dan Janzen does his usual forceful job of asserting South-North perspectives: the South possesses the great bulk of genetic resources, the North possesses the technology to exploit them, and the two are hardly talking about their joint interests. This theme is further examined by Cary Fowler with emphasis on biotechnology and genetic patenting, while Anil Agarwal presents an Indian viewpoint of social and

political factors. Michael Soule and Scott Mills consider conservation genetics writ large. Anthony Brown and Daniel Schoon assess plant population genetics. Several writers examine the promise and the problems of gene banks among other forms of *ex situ* conservation. In the final chapter, Arne Naess provides a stirring proclamation of the role of biodiversity in sustainable development, and explores the integrative approach required. The book concludes with three sets of recommendations directed at policymakers, administrators and managers, and researchers; the recommendations, however, are not very illuminating, being all too familiar and predictable.

Much of this might evoke the response: "So what else is new?" The principal value of the book lies in its focus on the genetic dimension of the biotic crisis, expressed through an in-depth treatment of questions such as

gene diversity (not at all the same as is usually understood by biodiversity), both inter- and intra-population variability, quantitative traits and genetic stasis, drift and erosion — all tackled within the context of conservation biology. How I wish I could have seen a book of this scope several years ago — and from this arises my chief criticism. The book presents papers from a conference held almost two years ago. Since then, we have probably lost well over 100,000 species, plus untold amounts of genetic variability. The conservation resource in shortest supply is not scientific expertise or funds, it is time. □

Norman Myers is at Upper Meadow, Old Road, Headington, Oxford OX3 8SZ, UK.

Sensing the potential

Gordon Malan

Biosensor Principles and Applications. Edited by Loïc J. Blum and Pierre R. Coulet. Marcel Dekker: 1991. Pp. 357. \$125.

Techniques and Instrumentation in Analytical Chemistry 11: Biosensors. By Frieder Scheller and Florian Schubert. Elsevier: 1992. Pp. 359. Dfl 315, \$161.50.

Immunochemical Assays and Biosensor Technology for the 1990s. Edited by R. M. Nakamura, Y. Kasahara and G. A. Rechnitz. American Society for Microbiology: 1992. Pp. 411. \$43 (members), \$51 (non-members).

BIOSENSORS are portrayed in scientific and popular journals as having immense potential in medicine, clinical chemistry, environmental monitoring and defence. But the image of a biosensor as a miniaturized probe or hand-held 'black box' that provides instantaneous analytical results is, with one or two notable exceptions, still largely science fiction.

Blum and Coulet have put together a useful book containing extensive reviews by international experts. Enzymatic amperometric and potentiometric biosensors are discussed, as well as new areas such as field-effect transistors (FETs) and luminescent and fibre-optic biosensors. Numerous summary tables are provided that compare technologies, biosensor properties and their sensitivities. A valuable penultimate chapter by P. Vadgama and M. A. Desai addresses the treatment of surfaces of biosensors for *in vivo* biocompatibility.

Scheller and Schubert's revision of *Biosensoren*, first published in 1989, presents a rather dated image, despite the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Rainforest in Brunei — loss of unique habitats can result in severe biodepletion.

vor species are likely to have lost many of their races and populations, so their gene reservoirs will have been grossly depleted. This will reduce their capacity to respond to new environmental pressures in a greatly changed biosphere: and in turn, this will probably bring about ecological instability of many different sorts.

True, there is much written about the adverse economic repercussions of genetic decline. There will be all the less potential for germplasm materials to supply new medicine, foods and industrial raw materials. But this immediate loss may prove to be less significant for human welfare in the long run than the broader consequence of draining the planetary gene pool. Regrettably, the second dimension of the biodiversity issue (or rather, biodepletion) is rarely mentioned, let alone discussed in detail (with the notable exception of writings by Michael Soule, such as *Viable*

addition of 20 more recent references. In their introduction, the authors deal cursorily with fundamentals, such as enzyme kinetics and equation-dominated mathematical modelling of enzyme amperometric electrodes. The main part of the book describes 'metabolism sensors', principally enzymatic electrochemical biosensors. The chapter entitled "Affinity sensors" introduces an artificial distinction because enzymatic electrochemical-detection of antibody-antigen reactions comprises most of this section: there is passing mention of recent developments using FETs, piezoelectric biosensors, and only a few of the many types of optical immunosensors.

Of the three books, I most enjoyed the crisp, concise and informative style of the American Society for Microbiology publication which reviews immunoassays thoroughly in the first two sections, and biosensors in the final third. The editors have provided a uniform style of presentation that makes it easy to forgive the use of the terms 'singlicate' (meaning singleton) and 'laboratorian' in one chapter. There are plenty of helpful drawings, diagrams, pictures and tables. In general, the authors refer only to the most important or relevant work in the field thus making the text useful to the student and experienced research worker alike. In addition to the chapters on the familiar enzymatic electrochemical biosensors, there are two interesting chapters on chemoreceptor-based and pharmacological sensors.

It is unfortunate that none of these books includes serious discussion of recent biosensor developments and, in particular, optical biosensors: these rely on evanescent field effects and have potential detection limits in the picomolar concentration range. The array of authors shows that development of enzymatic amperometric or potentiometric biosensors occurs mainly in academic laboratories. Nevertheless, a few commercial products are now available for measuring glucose, lactate, certain gases and various ions. The reason for the relatively slow commercialization of these methods since the first biosensor publication in 1962 by L. C. Clark and C. Lyons (*Ann. N.Y. Acad. Sci.* **102**, 29-45; 1962) is touched on in different ways; for example, chemical attachment to and stability of the enzyme at the sensor surface, nonspecific effects or fouling of the sensor surface, and detection limits in the micromolar range, all affect biosensor use. Thirty years on, the conclusion of most of the authors is that enzymatic electrochemical biosensors still 'have enormous potential'. □

Gordon Malan is at Amersham International p.l.c., Amersham Place, Little Chalfont, Buckinghamshire HP7 9NA, UK.



'Polar bears play-fighting' is taken from the cover of *An Introduction to Animal Behaviour* by Aubrey Manning and Marian Stamp Dawkins. This has been revised to include the many changes in the 12 years since the last edition. Published by Cambridge University Press, £40.00, \$69.95 (hbk), £16.95, \$24.95 (pbk).

Moscow, physics and beyond

John M. Charap

I. E. Tamm: Selected Papers. Edited by B. M. Bolotovskii and V. Ya Frenkel. Springer: 1991. Pp. 325. \$98, £49.50.

IGOR Tamm was one of the generation of Russian theoretical physicists who took what was a moribund and backward subject in their country to its present position of international prominence. From 1934 until his death in 1971 he was head of the theoretical physics department of the Lebedev Physical Institute of the Academy of Sciences of the USSR, where he established a lively research school inspired by his abilities as a teacher and his own wide-ranging research. His opposite number at the academy's Institute of Physical Problems was his younger contemporary Lev Landau; under their leadership, Moscow was to become an international centre of excellence for theoretical physics.

Tamm is best known for the work for which in 1958 he shared the Nobel prize with Ilya Frank. This was their interpretation of the Cerenkov effect, the radiation emitted by a charged particle when its speed exceeds the speed of light in the medium through which it is moving. The book includes three versions of this: the first brief paper with Frank, Tamm's later elaboration and extension

of that work, and his charming, elegant Nobel lecture, which includes generalizations to Mach waves in acoustics and to plasma waves.

Plasma physics had occupied his attention a decade earlier, for Tamm and his student Andrei Sakharov had been working on the theory of a magnetic thermonuclear reactor — the tokamak. Tamm's two classic papers from 1951, which together with one by Sakharov comprise the trilogy that established the feasibility of the tokamak, are included in this collection. For all their historic significance, one wonders what new readership they will now attract; they are, after all, reprints from an English translation published in 1960.

It is a pity that some of the other papers included in the selection are not translated, but reprinted in their original language — German, not Russian. In particular, Tamm's interpretation of the Raman effect in crystals — the scattering of photons from phonons — is illuminating and deserves to be made more accessible. It is also remarkable how early (1929) this mature application of quantum mechanical ideas was written. One is made aware of the resentment caused when no mention was made in Raman's Nobel-prize citation of the work of the simultaneous and independent investigations of Leonid Mandelstam and G. Landsberg. Mandelstam had been Tamm's teacher in Odessa, and Tamm had a lifelong admiration and affection for him. They collaborated on an interesting analysis of the uncertainty

relation between energy and time, also reprinted in this collection.

Tamm's work in 1932 on the electron energy levels at the surface of a crystal was, at the time, of only theoretical interest, and of no practical consequence. That has now changed, and these free nonsaturated valences of the surface atoms are important in point-contact phenomena, epitaxy and catalysis. He also worked on problems in nuclear physics and quantum field theory.

A few weeks after Enrico Fermi had published his theory of β -decay, Tamm and D. Iwanenko studied the possibility that the neutron-proton exchange interaction might be mediated by the exchange of a neutrino-electron pair, an idea he promptly rejected when he showed that the resulting interaction was much too feeble.

Tamm developed an approximation method for calculating meson-nucleon interactions, a difficult problem because the interaction, unlike that of electrodynamics, is strong and perturbation theory fails. The Tamm-Dancoff method (so-called because it had been independently derived by S. M. Dancoff, albeit some five years later; in fact, it is closely related to an idea of V. A. Fock from 1934) uses an expansion in the number of virtual particles rather than in the coupling constant. It has since been superseded, but attracted considerable attention in the 1950s.

The book also includes several of Tamm's appreciations of the lives and works of other physicists, notably an early analysis of the basic ideas underlying Faraday's research on electricity and magnetism. As Rudolf Peierls tells us in his preface, Tamm was a person of absolute integrity. He was outspoken in defence of the ideas of modern physics at a time when Marxists and Nazis (for their different reasons) reviled them. He was also an active participant in the Pugwash conferences, and can be credited with the idea of 'black boxes', unmanned seismic recorders to detect underground nuclear tests and so dispense with the dangers of on-site inspection and surveillance.

For all its virtues, I am still hesitant to recommend this book. The conjunction of reprints of Tamm's technical papers, some untranslated, with a selection from his *obiter dicta* is somehow unbalanced and unsatisfying. It is less than a complete guide to his scientific works, less than a scientific biography, less than an annotated bibliography. I do not know what the editors originally set out to achieve, but my guess is that they have failed. □

John M. Charap is in the Department of Physics, Queen Mary and Westfield College, Mile End Road, London E1 4NS, UK.

Listening post

Walter Wilczynski

The Evolutionary Biology of Hearing. Edited by Douglas B. Webster, Richard R. Fay and Arthur N. Popper. Springer: 1992. Pp. 859. \$129, £88.50.

A DISTINCTION can be made between comparative biology and evolutionary biology. Comparative biology describes the patterns of traits across the biome and identifies similarities and differences among organisms. It is often a necessary antecedent to evolutionary biology, and many comparative biologists move one step beyond the descriptions to suggest the transformations leading to diversity in biological systems. The transition is gradual, but comes finally when one shifts from asking what traits and transformations exist to where, why and how they arise.

The editors of this book declare their intention to orientate the volume towards the evolutionary biology of the auditory system, and have been remarkably successful in shepherding a large and diverse flock of contributors across the boundary from comparative description to evolutionary discussion. Their success means that this large, expansive book goes well beyond a more customary march from tympanic membrane to telencephalon, in which contributors describe at each stage the characteristics of their particular organism. There are elements of that here, and indeed the chapters are organized so that invertebrate systems are discussed first, mammals last, and other vertebrates between them in the characteristic order seen in books on comparative neurobiology. Moreover, the chapters do contain a substantial amount of comparative information, particularly about the peripheral auditory system and the related lateral line systems. But in addition to these interesting surveys of auditory system structure and function, the information in virtually every chapter is used to illuminate the evolution of a particular system or of hearing in general.

The book is strongly orientated towards the interrelated areas of neuroanatomy, neurophysiology and functional morphology of the central auditory pathways and ear. There are also chapters on behaviour, but they are in the minority. Some chapters, such as that of D. R. Ketten on the ears of marine animals, are narrowly focused, whereas others, such as the one by R. R. Fay on sound discrimination in vertebrates, take a more sweeping view. Some contributors provide detailed descriptions of a particular part of the

auditory system, as do B. L. Roberts and G. E. Meredith in their chapter on the efferent innervation of the ear. Others, while remaining anchored in the auditory system, take a more global approach by discussing the problem of understanding the evolution of hearing and of neural systems in general. For example, R. G. Northcutt looks at the relationship of ontogeny and phylogeny in neural evolution, whereas W. C. Stebbins and M. S. Sommers consider how to investigate the evolution of perceptual and behavioural processes.

J. R. Bolt and E. R. Lombard, who contribute one of four chapters covering the palaeontological data on auditory evolution, combine these approaches by presenting a good, clear survey of the fossil record of the amphibian stapes with cogent remarks on the interpretation of fossil evidence in general. Nevertheless, no matter how narrow or broad the focus, or how specific or general the outlook, all the authors manage to keep an evolutionary perspective on the comparative data they review.

Evolutionary questions take various forms. For example, A. Michelsen, T. E. Hetherington, and W. Plassmann and K. Brandle explain how the physics of sound and hearing organs constrain and shape audition in different organisms, whereas E. R. Lewis discusses how the fundamental dictates of signal processing may have led to the remarkable convergence in the physiological properties of auditory receptor organs in vertebrates. P. M. Narins, H. Römer, and N. A. M. Schellart and A. N. Popper concern themselves more with environmental or other selective pressures shaping the structure and function of auditory systems.

Still others focus on phylogenetics, such as M. R. Miller, who uses a review of auditory traits in lizards to construct and test a phylogeny of this vertebrate group, or C. A. McCormick and P. Fritsch, both of whom use comparative data to speculate on the antecedents of terrestrial hearing. These general areas — physical constraints, selection and phylogenetics — are essential for understanding the evolution of any biological system.

This huge book is at times understandably unwieldy, and certainly not one for the casual reader uninterested in the details of auditory structure and function, but it is nevertheless valuable. By moving beyond comparative biology to evolutionary biology, the editors and authors have constructed a most interesting and useful book. □

Walter Wilczynski is in the Department of Psychology, University of Texas, Austin, Texas 78712, USA.

Molecular replication

Leslie E. Orgel

Simple replicating molecules are likely to have played a critical role in the origin of life. Recent experiments show that non-enzymatic replication is conceivable in a wide range of synthetic chemical systems. The challenge now facing these studies is how to develop information-coding systems from simple prebiotic precursors.

THE double-helical structure of DNA and an outline of the mechanism of nucleic acid replication were established before the growth of interest in molecular replication as an area of organic chemistry. Consequently, early experiments on molecular replication used substrates closely related to polynucleotides¹⁻³. Organic chemists later began experimental work on replication in a more general context, as an aspect of self-organization⁴⁻⁹, and research on prebiotic chemistry has begun to focus on the possible role of simpler replicating polymers as precursors of nucleic acids¹⁰. My main purpose here is to classify non-enzymatic replicating systems and to examine their relevance to the problem of the origin of life.

All replicating systems are, by definition, autocatalytic and all autocatalytic systems result, in some sense, in replication. The relevance of a model system to the origin of life depends on the nature of the molecule that replicates, the mechanism of replication, and the conditions that are required to permit replication. Two very different types of replicating system must be considered. The first does not involve direct replication of a polymeric molecule, but may help to create an environment favourable for such replication. Systems involving the autocatalytic synthesis of one of the building blocks of a replicating polymer, or the components of a membrane, might be of this kind. I refer to such systems as non-informational autocatalytic systems. The second type of system involves the replication of informational polymers made up from two or more different monomers. Here, the conservation of sequence information is paramount, so I refer to such systems as informational replicating systems.

This discussion is mainly concerned with non-enzymatic informational replicating systems. No such system has been realized in the laboratory, but two classes of model system that embody important aspects of replication have been described in detail. In the first, non-informational molecules that can be considered as simplified representatives of an informational macromolecule have been shown, unequivocally, to replicate. A hexanucleotide of defined sequence, for example, has been shown to replicate using the 5'-terminal and 3'-terminal trinucleotides as building blocks¹. In the second, informational polynucleotides have been copied, that is sequence information has been transmitted in a chemical reaction, but replication has not been achieved. The sequence CCGCC, for example, has been shown to catalyse the synthesis of GGCGG from a mixture of activated G and C nucleotides¹¹. But the reciprocal reaction that is needed to complete a replication cycle has not been demonstrated. The combination of replication with conservation of sequence information is the holy grail of the field.

The interest of replicating systems goes beyond the light that they may throw on the origins of life on the Earth. Perhaps work on replication will ultimately lead to a new kind of chemistry of organic polymers in which the concepts of natural selection and evolution will play a part⁴⁻⁹. Although this is not the main focus of this analysis, much of what is said will be applicable to replicating polymer systems of any kind.

Non-informational autocatalytic systems

Many autocatalytic systems have been examined by physical and physical organic chemists, but most of them have little relevance to our subject. A few classes deserve mention. An attractive hypothesis suggests that particular polymeric products formed from a reactive molecule such as formaldehyde or hydrogen cyanide can specifically catalyse (or photocatalyse) the formation of further molecules of the same type. It would be exciting, for example, if the addition of ribose to a polymerizing solution of formaldehyde led to the production of additional amounts of ribose at the expense of other sugars. Unfortunately, the accumulated evidence indicates that preformed hydroxy-aldehydes and sugars accelerate the rate of polymerization of formaldehyde but do not increase the specificity of the reaction¹². Similarly, the products of the polymerization of hydrogen cyanide accelerate further polymerization but are not known to lead to a specific pattern of products. The possibility of specific autocatalysis in this kind of system certainly exists, but no example is known at present.

Autocatalysis has been discussed extensively as a possible mechanism for amplifying optical asymmetry on the primitive Earth. A number of primary mechanisms have been proposed that would lead to the development of a very slight excess of one enantiomer of an organic molecule (such as an α -amino acid) over the other¹³. Then autocatalytic synthetic or degradative processes are shown, theoretically, to amplify the enantiomeric excess until it reaches a significant level¹³. I am not persuaded that these mechanisms are either plausible or necessary. The formation of mixtures of crystals each of a single handedness from optically inactive solutions is well established¹⁴. Very recently an example has been reported in which nucleation of a supersaturated solution of sodium chlorate by the first-formed crystal leads to the separation of many crystals all of the same handedness¹⁵. These observations are not obviously related to the origins of life.

The production of membranes, micelles, vesicles and so on is clearly relevant to the origins of life, as they provide mechanisms for keeping together soluble molecules that have been involved in each other's synthesis. This is essential if natural selection is to work efficiently. There has been some theoretical discussion of autocatalysis in connection with the formation and function of membranes^{16,17}, but it is only recently that an illuminating experimental system has been described^{18,19}.

Reverse micelles consisting of microdroplets of aqueous LiCl solution form when a bulk aqueous solution of LiCl is shaken with 50 mM sodium octanoate in a 9:1 (v/v) isooctane/1-octanol mixture. The aqueous droplets are stabilized in the organic bulk solvent by the surfactant action of octanoic acid molecules, which orient themselves in the interface between the aqueous and hydrophobic environments. If the octyl ester of octanoic acid is added to the system, it co-localizes at the aqueous-organic interface, and is there hydrolysed by Li⁺ ions. One of the products of the reaction is octanoic acid. Thus pre-existing micelles stabilized by octanoic acid catalyse the formation of more octanoic acid which could be used to stabilize

more micelles. Clearly this particular replication system is unlikely to have operated on the primitive Earth. But the model is very suggestive, and may encourage work along the same lines using more plausibly prebiotic substrates.

Non-informational replicating systems are not genetic systems. They cannot evolve by natural selection, because they do not store information in a stable way. This is not to deny that such systems may have been important for the origins of life—they may well have been essential for the creation of an environment in which informational replication could get started.

Informational replicating systems

Nucleic acid replication provides a useful starting point for a discussion of informational replicating systems in general. A highly conceptualized model of the replication of double-stranded RNA using activated mononucleotides as substrates is shown in Fig. 1*a* and *b*. We distinguish the following steps in the cycle: (1) separation of a double-helix into its component strands; (2) initiation of replication by the formation of a dimer from two activated monomers at the terminus of each strand; (3) propagation of synthesis by successive addition of monomers until each strand is copied from end to end.

The operation of the cycle illustrated in Fig. 1*a* and *b*, without the help of enzymes, has not been achieved, although some of the subreactions have been reported (ref. 11, discussed later). Here, I want to concentrate on the essential features of the cycle, in order to see how to generalize it. For the initiation and elongation steps to work, the preformed template strand must be able to align one incoming activated monomer next to another, or at the end of a partially synthesized complementary chain, in such a way as to favour the formation of a new phosphodiester bond. In RNA replication the geometric constraints of Watson and Crick base pairing are assumed to guaran-

tee that the correct alignment will be achieved only if a correct base pair is formed. In the other main step in the cycle it is essential to separate the strands of the double helix from each other without breaking the backbone of either strand.

Figure 1*a* and *c* indicates the obvious generalization of this idea. A preformed chain made up by polymerizing the units U_1, U_2 and so on must be able to interact with these same monomeric subunits according to some well-defined pairing rule. The geometry of the complex formed by this interaction must aid the joining together of the subunits in a sequence matching that of the preformed chain. It must then be possible to separate the two strands without breaking the backbone of either strand. The last requirement is kinetic rather than thermodynamic. The interstrand interactions need not be weaker than the intrastrand interactions holding the backbone together, but there must be some way of dissociating the interstrand bonds without breaking the intrastrand bonds.

Apart from the requirements described above, few of the features of RNA replication are essential for a general replication model. There could be fewer or more than four subunits, pairing might occur between identical or complementary subunits, each substrate molecule might contain several monomeric units (trinucleotides might be used in an RNA system, for example), chains could run in parallel or antiparallel directions, and replication might occur in a homogeneous aqueous or non-aqueous solution or on a surface. The interaction between chains could be mediated by hydrogen bonding, electrostatic interaction, hydrophobic forces, covalent bonds, or co-ordination of functional groups on the two chains to common metal ions. Chain growth could occur in either direction, or possibly in both directions from an internal initiation site.

Even the stability of arbitrarily long double-stranded molecules is not essential, provided that a sufficiently extended structure is present to guarantee that a sequence of informational units of a preformed chain can pick out their partners and aid their assembly into a new chain. In this context, it is interesting that certain RNA polymerases such as Q β RNA polymerase produce single-stranded products; double-helical RNAs are not substrates for these polymerases^{20,21}.

What has been achieved?

Replication. The first successful demonstration of molecular replication¹ was strictly modelled on nucleic acid replication (Fig. 2). The two trinucleotide analogues **I** and **II** and the hexanucleotide analogue **III** were prepared by standard synthetic methods. The product formed by linking **I** and **II** by a phosphodiester bond is **III** (Fig. 2*a*). The sequences were chosen because they form the ternary complex **T** in which the 3'-phosphate of **I** is brought close to the 5'-hydroxyl of **II** (Fig. 2*b*). The addition of a water-soluble carbodiimide to the solution, therefore, efficiently joins together **I** and **II** to form **III**, and converts the ternary complex **T** to the binary complex **B**. Thus a molecule of **III** catalyses the production of another molecule of **III** from **I** and **II**.

This result, and related results using different oligonucleotide-related substrates that replicate more efficiently^{2,3,22,23}, have not unequivocally demonstrated the possibility of informational replication. But they strongly imply that at least limited informational replication is possible. First, it seems unlikely that the replication of **III** would be inhibited by the presence of oligonucleotides with sequences unrelated to **I**, **II** and **III** (ref. 2). Second, a chain of nine or twelve residues would almost certainly form a double-helical complex analogous to **B** but with three or four trinucleotides. The trinucleotides could then be ligated with an appropriate condensing agent. Thus replication of such polymers is at least likely. Ribozyme-catalysed copying of a polynucleotide using oligonucleotides as substrates has already been achieved²⁴.

The initial experiments using trinucleotides¹ revealed an important obstacle to replication. The increase in the rate of

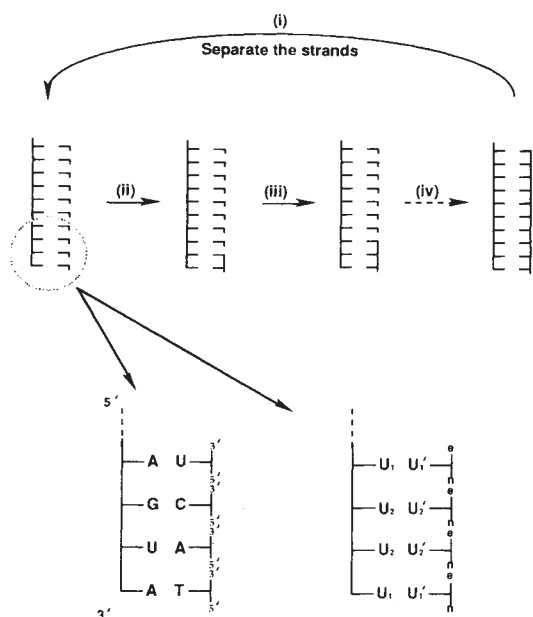


FIG. 1 Idealized scheme for informational replication via a double-stranded intermediate. *a* (i), The double-stranded product formed in the previous round of replication dissociates into single strands and monomers line up on them; (ii) a dimer is formed at the end of the template; (iii) this dimer acts as a primer for addition of another monomer; and (iv) the process continues until the template is filled. *b*, The template-monomer complex for RNA replication. The monomers are activated nucleoside-5'-phosphates. *c*, The template-monomer complex for a generalized template-directed reaction. The monomers are assumed to be bifunctional, involving nucleophilic (n) and electrophilic (e) groups that can react with each other to form a covalent bond. U_1 pairs with U'_1 and so on.

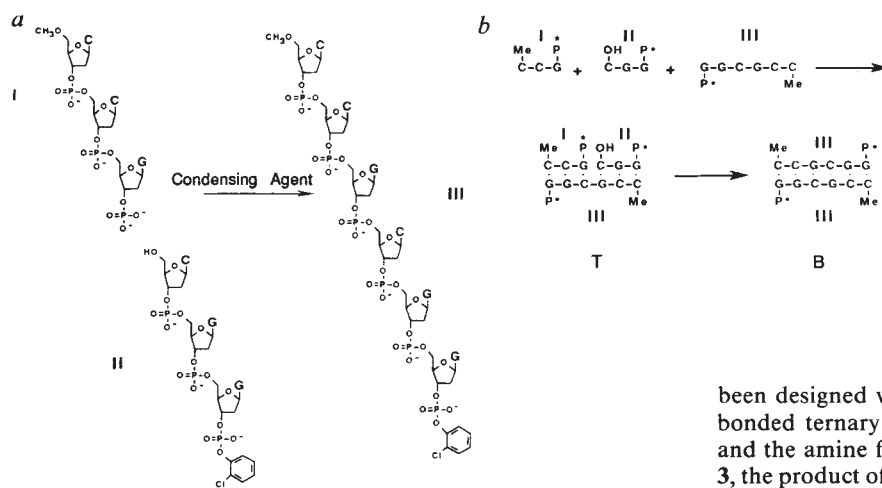


FIG. 2. Replication of a hexanucleotide. *a*, The fundamental reaction is the ligation of the two trinucleotide analogues I and II to form the hexanucleotide analogue III; *b*, The catalytic mechanism: a Watson-Crick double-helical ternary complex T is formed from one molecule each of I, II and III. Formation of a phosphodiester bond between adjacent 3'-phosphate and 5'-hydroxyl group converts T to the binary complex B made up from two hexamers. The binary complex B must dissociate into two molecules of III to initiate the next round of replication (see text). Adapted from ref. 1.

synthesis of III due to autocatalysis is proportional to the square root of the concentration of III present, rather than to the concentration of III. The reason is that most of the molecules of III present in the solution are tied up in unbroken double helices (B in Fig. 2b), but only isolated molecules can act as templates. The number of isolated molecules increases roughly as the square root of the total concentration. Oligomers containing more than six residues would form much more stable double helices than B, so autocatalytic synthesis would be expected to slow down before substantial concentrations of replication products had formed. To overcome this difficulty it will almost certainly be necessary to cycle the temperature of the reaction mixture, or to cycle some other variable that controls the stability of double-helical complexes relative to single strands.

Rebek and coworkers have approached the problem of self-replication differently⁴⁻⁹. Instead of working with molecules related to nucleotides in aqueous solution, they have synthesized relatively complicated organic substrates that replicate in non-aqueous solvents. The basis of their experiments⁶ can be understood by reference to Fig. 3. The subunits 1 and 2 contain an activated carboxyl group and a primary amine, respectively. When 1 and 2 react together an amide bond is generated and the molecule 3 is formed. The triad of molecules 1, 2 and 3 have

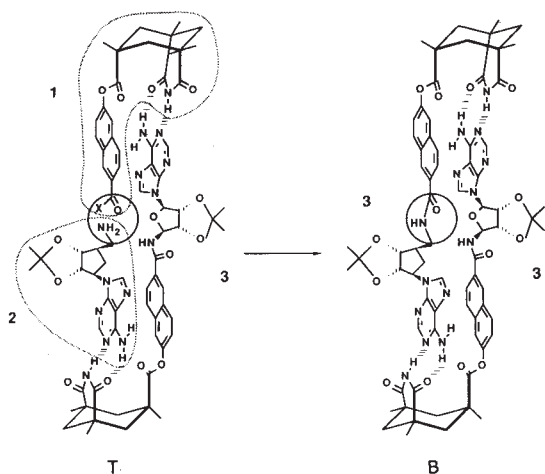


FIG. 3 Replication of novel hydrogen-bonding substrates. The trio of molecules 1, 2 and 3 have been designed ingeniously to form the hydrogen-bonded ternary complex T, in which an activated carboxylic acid is brought close to an amine group. (The relevant groups are enclosed in a circle.) When an amide bond is formed, T is converted to the binary complex B, a dimer of 3. Separation of B into two molecules of 3 makes possible another round of replication.

been designed with great ingenuity, for they form a hydrogen-bonded ternary complex T in which the carboxyl group of 1 and the amine function of 2 are held in close proximity. Thus 3, the product of the reaction between 1 and 2, aids the synthesis of further molecules of 3 from 1 and 2. This is exactly what is required for self-replication.

The kinetics of the autocatalytic synthesis of 3 from 1 and 2 demonstrates a roughly square-root dependence on the concentration of 3 (ref. 6). Again this is the result of the stability of dimer B formed by association of two molecules of 3. Rebek has extended his studies to other kinds of molecule that can form the basis for a replicating system and is beginning to study more-complicated systems in which more than one substrate of replication can be formed in the reaction mixture^{7,8}. He uses expressions such as "mutation" and "riotously fertile hybrid" to describe the outcome of these experiments. A discussion of this picturesque nomenclature is given in ref. 9.

The experiments described by Rebek and his coworkers do not address the problem of informational replicating systems directly. But like those of von Kiedrowski¹, they can almost certainly be generalized by arranging for several different substrate molecules to polymerize on a preformed linear template. This provides the exciting prospect of studying the replication of organic polymers very different from nucleic acids.

Terfort and von Kiedrowski have now developed another self-replicating system involving relatively simple organic molecules, derivatives of 2-formylphenoxyacetic acid 4 and 3-aminobenzamidine 5, which condense autocatalytically in dimethylsulphoxide to form anils 6 (Fig. 4)²⁵. The initial experiments suggest that, in some experiments, the rate of the autocatalytic reaction depends linearly on the concentration of 6. This is in sharp contrast to the square-root dependence on catalyst concentration in the reactions. Presumably the ternary complex T is in some cases more stable than the binary complex B. This new system suggests that simple molecules unrelated to

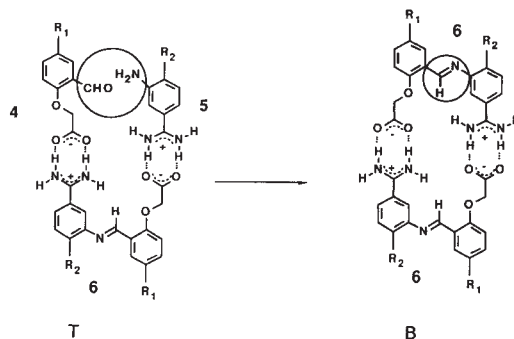


FIG. 4 Self-replication based on interaction between carboxylate anion and amidinium cation. The anils 6 form ternary complexes T with the carboxylate-containing ions 4 and the amidinium-containing ions 5 by means of hydrogen bonding and electrostatic interaction. Formation of a Schiff's base generates another molecule of 6 from 4 and 5 to give the binary complex B. The dissociation of B to two molecules of 6 begins a new round of replication.

nucleotides may provide the basis for an exponentially replicating system.

Copying. Some polynucleotides or polydeoxynucleotides will direct the synthesis of complementary oligonucleotides. I refer to this type of sequence-specific oligonucleotide synthesis as 'template-directed synthesis' (Fig. 5). Progress before about 1987 has already been reviewed¹¹, so only a summary of the results will be given here.

In describing how well a given template sequence is copied, one must specify the efficiency, the fidelity and the regio-specificity. The efficiency is measured by the proportion of template molecules that succeed in directing the synthesis of a complementary sequence, the fidelity by the probability that a complementary base rather than a noncomplementary base will be inserted opposite a given base in the template, and the regio-specificity by the ratio of natural 3'-5' linkages to unnatural 2'-5' linkages. The first main conclusion is that the template-directed reactions of most activated nucleotides are inefficient and nonregiospecific. A considerable search was required to find a set of activated nucleotides, the 5'-phosphoro(2-methyl)imidazolides, that undergo efficient and highly regiospecific template-directed reactions²⁶.

With these reagents, the following results have been established. (1) Poly(C) directs the synthesis of long oligo(G)s. The reaction is highly regiospecific and the fidelity is excellent. If poly(C) is incubated with a mixture of guanosine-5'-phosphoro(2-methyl)imidazolidine and the corresponding derivative of A, T or G, the second base makes up less than 1% of the product oligonucleotide. (2) Random copolymers containing 60% or more C residues direct the synthesis of products containing G and the complements of the other bases present in the template. The template poly(C, A, U), for example, aids the synthesis of products containing G, U and A²⁷. (3) Many defined oligonucleotide sequences are copied faithfully. These include sequences containing an excess of C residues and A, T and G residues isolated from each other by at least three C residues. Runs of G residues are also copied into runs of C residues if the formation of self-structures by G residues can be avoided²⁸. The sequence 5'-CCCGCCCGCCGCC-3' is the longest oligomer that has been copied successfully at present. The yield of full-length products in this case never exceeds 2%, but the yield including truncated products can amount to 50% or more²⁹. (4) We are unlikely to find a pair of complementary sequences each of which facilitates the synthesis of the other using nucleoside-5'-phosphoro(2-methyl)imidazolides as substrates. The chief obstacles are the formation of intermolecular complexes involving a tetrahelix formed by sequences of consecutive G residues³⁰, and the formation of intramolecular complexes in which the

molecules fold back on themselves and form Watson-Crick double-helical segments.

In summary, copying with information transfer can be achieved in this relatively simple system, but there are severe obstacles to replication. Some of the obstacles may be attributable to the choice of reagents and reaction conditions, but others seem to be intrinsic to oligonucleotide replication.

Template-directed vinyl polymerizations. Vinyl homo- and copolymers are of great commercial importance, so the polymerization of simple vinyl monomers has been explored in detail. Here I can only discuss a few examples and cite a very few key references.

The polymerization of methyl methacrylate has been studied extensively³¹ and methods for synthesizing the isotactic (all asymmetric carbon atoms having the same optical configuration) and syndiotactic (alternating *R* and *S* centres) polymers have been developed. If the solutions of the isotactic and syndiotactic polymers, for example in dimethylformamide, are mixed, a so-called stereocomplex containing two parts of the syndiotactic to one part of the isotactic polymer is formed. If methyl methacrylate in dimethylformamide solution is polymerized in the presence of isotactic poly(methyl methacrylate), polymerization is accelerated and the syndiotactic polymer is formed. Conversely, if methyl methacrylate is polymerized in the presence of the syndiotactic polymer, the syntactic product is formed. In either case, the final product isolated from the solution is the stereocomplex (see for example Fig. 6).

These reactions have many of the properties required for non-informational replication. If the stereocomplex could be separated into its isotactic and syndiotactic components, each could be used to initiate another 'round of replication'. The detailed kinetics of the reaction are strikingly reminiscent of the non-enzymatic copying of polynucleotides: a short 'primer', for example, is required under many conditions.

Several systems have also been described³¹ in which homopolymers such as poly(2-vinyl pyridine), poly(ethylene imine) and poly(*N*-vinyl pyrrolidone) act as templates to accelerate the polymerization of unrelated monomers such as acrylic acid or phenylalanine-*N*-carboxyanhydride. The polymerization of positively charged *N*-vinyl imidazole on a polyanionic poly(methacrylic acid) template in aqueous solution has been studied in great detail³². Substantial rate accelerations have been reported, and the effect of the template on the stereospecificity of the reaction has been discussed.

The relevance of these reactions to informational replication should not be exaggerated. They are all concerned only with the copying of a homopolymer, and none addresses the problem of chain separation and exponential growth. The most impressive rate accelerations in aqueous solution occur in systems involving oppositely charged polymer and monomer components, where the interaction is electrostatic and therefore may not be very specific. Nonetheless, it is striking that 'copying' occurs readily in vinyl and in other less-studied types of polymerization^{33,34}. It encourages the belief that 'copying' is a rather general phenomenon.

Difficulties for non-enzymatic replication

The simplicity and generality of the cycle illustrated in Fig. 1a and c suggests, at first sight, that the design of a novel informational replicating system should be a simple matter. All that is needed is an elaboration of one of the schemes already used for self-replication or of some similar scheme. Each subunit must carry both of the functionalities required to form the backbone bond, rather than just one of them (Fig. 1c). In addition, something similar to the geometrical equivalence of the base pairs must be designed into the scheme. Each subunit in the template chain must pick out a unique partner for insertion in the newly synthesized chain. The geometry of the interaction between the preformed chain and the activated subunits involved in replication must place the bond-forming functionalities in

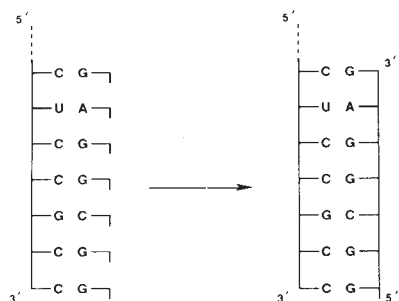


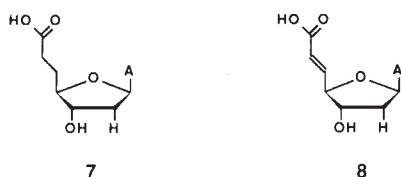
FIG. 5 Template-directed synthesis. A preformed oligonucleotide chain forms a double-helical complex with complementary activated nucleoside-5'-phosphates. The geometry of the helix brings the 5'-phosphate of each monomer close to the 3'-hydroxyl of its neighbour, thus facilitating the formation of a phosphodiester bond. The outcome is the formation of an oligomer complementary to the template. This model is greatly oversimplified. The experimental results and their interpretation are reviewed in ref. 11.

positions that favour reaction independently of the nature of the particular pair of subunits that are involved. It all sounds simple in theory; it may or may not turn out to be simple in practice.

Understanding how replication could have occurred on the primitive Earth involves further constraints. The subunits of the self-replicating polymer must be potentially prebiotic molecules, that is molecules that could have accumulated on the Earth without the intervention of organic chemists. Different authors will draw the line between plausibly and implausibly prebiotic molecules in different ways. Most would accept amino acids as prebiotic and reject the molecules illustrated in Figs 3 and 4. Whether ribonucleotides are prebiotic molecules is by no means clear^{35,36}. However generously 'prebiotic' is interpreted, it is obvious that it will be harder to realize a replicating system if we are restricted to prebiotic molecules and cannot use molecules specially designed for replication.

Finally, another level of difficulty needs to be recognized when we think about the origins of life. The subunits of an informational polymer are likely to be trifunctional—two functionalities are needed to hold the backbone together, and a third is needed to provide for specific interchain interaction. The prebiotic production of a particular trifunctional molecule may well be accompanied by the production of numerous isomers and closely related molecules. Because they have structures related to that of the subunits of the replicating polymer, these 'byproducts' would probably inhibit replication. The inhibition of the template reactions of monomers of one optical handedness by monomers of the other handedness is likely to be particularly troublesome³⁷.

If we suppose that it is possible to solve the nontrivial problem of designing subunits that interact to form a macromolecular complex of the type suggested in Fig. 1, there are still difficulties that need to be overcome to make efficient replication possible. These difficulties severely constrain the types of subunit that are acceptable. If all subunits are of the same category, they must contain both of the functionalities that interact to form the backbone bond. In general, this leads to the possibility of efficient internal cyclization reactions if the repeating unit of the backbone contains a chain of five or six atoms³⁸, and reasonably efficient cyclization for longer side-chains³⁹. This problem can be overcome either by using subunits too small to cyclize, such as α - or β -amino acids but not γ -amino acids, by using two categories of subunit, for example diamines and dicarboxylic acids, or by designing subunits that cyclize with difficulty or not at all. The nucleotide analogue **7** cyclizes so efficiently that it does not polymerize on a poly(C) template, but the analogue **8** cannot cyclize and so can be made to polymerize on the template⁴⁰. It may be significant that nucleotides derived from ribose cyclize less readily than most other nucleotides and nucleotide analogues⁴¹.



Even if direct cyclization of the monomer can be avoided as with α -amino acids, there is still the danger that short oligomers will cyclize. The formation of diketopiperazines (cyclic dipeptides), for example, reduces the efficiency of many direct polymerizations of α -amino acids. Diketopiperazine formation from dipeptidyl esters must have been a considerable obstacle to the evolution of protein synthesis. The initiation of polypeptide synthesis with *N*-formylmethionine in eubacteria

may have originated as a protection against diketopiperazine formation⁴².

In addition to difficulties with the choice of substrates, there are others that arise from the properties of the polymeric products. The most obvious is the difficulty of separating a newly formed strand from the template on which it was synthesized. Optimally, monomers should associate as strongly as possible with the preformed template and react as efficiently as possible, but the product oligomer should be separated from the template as easily as possible. Paradoxically, a poor fit between template and product strands may be helpful (Fig. 7), if there is a good local fit between the structure of the transition state for bond formation and the structure of the template.

A double-helical nucleic acid could easily be dissociated into single strands, for example by heating. It is more difficult to provide plausible mechanisms for preventing the strands from re-annealing when the temperature is lowered as it must be to permit a new round of template-directed synthesis from monomers. Segregation of the complementary strands by adsorption on a mineral surface is perhaps the most straightforward solution. A few experiments in my laboratory using hydroxylapatite as adsorbent and poly(C):poly(G) as the double-stranded nucleic acid were unsuccessful, but further work along these lines seems justified.

If the segregation approach is unsuccessful, the solution to the problem of polynucleotide replication will require considerable departures from the simple model suggested in Fig. 1. Some possibilities are discussed later. As we have seen, in the more general context of alternative genetic systems, the problem of strand separation may be less severe if, because of the structure of the backbone, the double-stranded structure is less stable than in nucleic acids.

Internal self-structure, formed when the template folds back on itself and forms double-helical stems, is another major obstacle to replication of polynucleotides, because monomeric nucleotides cannot disrupt the internal self-structure. Self-structure is always likely to be a serious problem if template and substrate strands each include all possible residues⁴³. It could, however, be avoided by assigning residues so that no two residues in any chain interact strongly with each other. A polynucleotide system based on copolymers of A and C and copolymers of G and U would escape the self-structure problem. The problem could be avoided in a similar way in other copolymers, for example by using strands that contain positively or negatively charged side-chains, but not both.

Finally, experience with polynucleotides suggests that co-optimization of the kinetics of incorporation of different subunits will be difficult⁴⁴. In the case of polynucleotides, the detailed structure of double-helical complexes is sequence-dependent. The conformation of the template-primer-monomer complex is likely to be even more variable. Consequently, an activated derivative that has been chosen to optimize one type of addition, say of one G residue to another, is not optimal for a different addition, say of a U residue to a C residue. Similar problems may be expected in other replicating systems.

Specificity, fidelity and mutation

The copying of nucleic acids is achieved with remarkably few errors. The simplest RNA polymerases incorporate an inappropriate base about once in 10^3 – 10^4 bases. DNA polymerases make use of elaborate error-correction mechanisms that may cut the error rate to below 10^{-8} . Many mutations are caused by copying errors; this component of the mutation rate is one measure of the lack of specificity of the copying enzyme. At least from the chemist's point of view, a very high mutation rate indicates an inadequately faithful copying mechanism.

The non-enzymatic copying of a poly(C) template to give oligo(G) products has an error rate of less than 1%. Recent results suggest, however, that the incorrect incorporation of G residues, particularly opposite T (and presumably U) residues

in the template, is considerably higher (ref. 28, and T. Wu and L.E.O., unpublished results). There are theoretical reasons for believing that natural selection cannot operate efficiently if the mutation rate is much greater than the reciprocal of the genome length (number of bases)⁴⁵. Clearly the challenge for organic chemistry is not to introduce mutations, but to lower the error rate in non-enzymatic systems to values comparable to those achieved by enzymes.

Alternative models

A number of models have been proposed in which some closed cycle of chemical reactions evolves to great organized complexity without any need for residue-by-residue copying of informational polymers⁴⁶⁻⁴⁸. It is usually assumed that some of the products that are formed in a spontaneous polymerization reaction are specific catalysts for the synthesis of other products, and that a small subset of the products can be chosen so that the synthesis of each member of the subset is catalysed specifically by one (or more) other members of the subset. It is always difficult in such theoretical models to see how to close the cycle without making unreasonable assumptions about the specificity of catalysis.

There are other, less radical, alternatives to the simple model illustrated in Fig. 1 that retain the central role of residue-by-residue copying. I have already mentioned the possibility that the extended double-strand form of the replicated molecule might be unstable, so that the new strand would peel off the template as it formed (Fig. 7). This would eliminate the need to separate the strands by cycling the temperature, pH or ionic strength of the environment (for a detailed discussion, see ref. 49).

Exponential replication by repeated chain extension, followed after an appropriate number of steps by chain breakage, is another possibility. This mode of replication avoids initiation of new chains, which is more difficult than the extension of primers. Such mechanisms are probably inapplicable to non-enzymatic synthesis of RNA, because spontaneous hydrolysis generates 2'- or 3'-phosphates that are not substrates for chain extension. It might be a useful mechanism in other polymer systems or in an RNA world, after a ribozyme evolved to cut RNA leaving a product terminated by a 5'-phosphate.

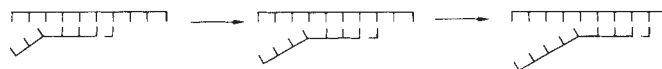


FIG. 7 Replication without the formation of a stable double-stranded product. Here we assume that a double strand is stable only over a length of three units. As each subunit is added to the right, one unit of the polymer peels off at the left.

Finally, I should mention the ingenious theories of Cairns-Smith who suggests that the first informational replicating system was inorganic⁵⁰. He claims, for example, that the pattern of lattice defects or cation substitutions in a preformed layer of a clay mineral such as montmorillonite can induce a similar pattern in a newly formed layer: the entity that replicates is the pattern of defects or cations. I am unaware of any example of replication or of stereospecific catalysis by clay minerals as envisaged by Cairns-Smith. In the absence of a detailed and convincing structural model of replication (or experimental evidence), I am not persuaded that patterns of defects or cation substitutions that are complicated enough to catalyse stereospecific organic reactions are capable of replication.

Prebiotic informational polymers

An extended discussion of the chemical nature of potentially prebiotic replicating systems⁵¹ would be premature. Briefly, hydroxy acids⁵², amino acids (ref. 53, and A. Rich and X. Zhang, personal communication), phosphomonoesters of polyhydric alcohols⁵⁴, aminoaldehydes⁵⁵ and molecules containing two sulphhydryl groups⁵⁶ have all been considered as simple prebiotic monomers from which an informational polymer could be constructed. A comprehensive list would no doubt contain many additional entries including copolymers of formaldehyde with amines, urea and so on. In one scheme the D- and L-enantiomers of a single substance, glyceric acid, serve as the two independent informational subunits of a replicating copolymer⁵².

The interaction between strands might be mediated by hydrogen bonding, electrostatic interaction in association with hydrogen bonding as in the interaction of carboxylate groups with amidine or guanidine cations, covalent bonding as between alcohols and carboxylic acids, or possibly by hydrophobic interactions. A slightly more complicated interaction in which metal ions hold chains together by binding simultaneously to anionic side-groups in each chain is also possible.

Prospects

The next few years should see the beginning of an experimental search for novel informational replicating systems. The most sophisticated approaches to molecular design will no doubt be used by the organic chemistry community. The remarkable success of a computer-assisted search for a novel 'antisense' oligonucleotide analogue with a peptide-like backbone⁵⁷ may indicate that prediction of stable complementary structures, at least, is now possible. Those prebiotic chemists who believe that some simpler informational polymer preceded RNA may find it harder to design complementary structures using the limited arsenal of monomers at their disposal. But modern structural methods should help at least to exclude many unstable structures. It seems likely that informational replication will be achieved in the next decade, and that it will throw new light on the origins of life. □

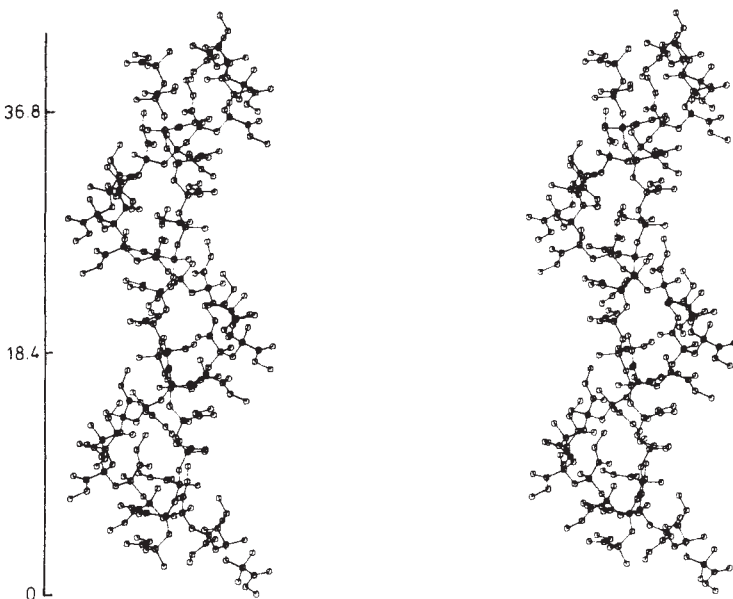


FIG. 6 Stereoview of the stereocomplex of isotactic and syndiotactic poly(methyl methacrylate) (distances are given in Å).

Leslie E. Orgel is at the Salk Institute for Biological Studies, PO Box 85800, San Diego, California 92186-5800, USA.

1. von Kiedrowski, G. *Angew. Chem. int. Ed. Engl.* **25**, 932-935 (1986).
2. von Kiedrowski, G., Wlotzka, B. & Helbing, J. *Angew. Chem. int. Ed. Engl.* **28**, 1235-1237 (1989).
3. von Kiedrowski, G., Wlotzka, B., Helbing, J., Matzen, M. & Jordan, S. *Angew. Chem. int. Ed. Engl.* **30**, 423-426 (1991).
4. Tjivikua, T., Ballester, P. & Rebek Jr, J. *J. Am. chem. Soc.* **112**, 1249-1250 (1990).
5. Rotello, V., Hong, J.-I. & Rebek Jr, J. *J. Am. chem. Soc.* **113**, 9422-9423 (1991).
6. Nowick, J. S., Feng, Q., Tjivikua, T., Ballester, P. & Rebek Jr, J. *J. Am. chem. Soc.* **113**, 8831-8839 (1991).
7. Rebek Jr, J. *J. Chem. Ind.* **1992**(3), 171-174 (1992).
8. Hong, J.-I., Feng, Q., Rotello, V. & Rebek Jr, J. *Science* **255**, 848-850 (1992).
9. Feng, Q., Park, T. K. & Rebek Jr, J. *Science* (in the press).
10. Joyce, G. F., Schwartz, A. W., Miller, S. L. & Orgel, L. E. *Proc. Natn. Acad. Sci. USA* **84**, 4398-4402 (1987).
11. Joyce, G. F. in *Cold Spring Harbor Symp. Quantitative Biology*, Vol. LII, 41-51 (Cold Spring Harbor Press, New York, 1987).
12. Mizuno, T. & Weiss, A. H. in *Advances in Carbohydrate Chemistry and Biochemistry* (eds Tipson, R. S. & Horton, D.) 173-227 (Academic, New York, 1974).
13. Bonner, W. A. *Origins Life* **21**, 59-111 (1991).
14. Jacques, J., Collet, A. & Wilen, S. H. *Enantiomers, Racemates, and Resolutions* (Krieger, Malabar, Florida, 1991).
15. Kondepudi, D. K., Kaufman, R. J. & Singh, N. *Science* **250**, 975-976 (1990).
16. Morowitz, J. J., Heinz, B. & Deamer, D. W. *Origins Life* **18**, 281-287 (1988).
17. Luisi, P. L. & Varela, F. J. *Origins Life* **19**, 633-643 (1989).
18. Bachmann, P. A., Walde, P., Luisi, P. L. & Lang, J. *J. Am. chem. Soc.* **112**, 8200-8201 (1990).
19. Bachmann, P. A., Luisi, P. L. & Lang, J. *Nature* **357**, 57-59 (1992).
20. Spiegelman, S. Q. *Rev. Biophys.* **4**, 213-253 (1971).
21. Blebricher, C. K. in *Evolutionary Biology*, Vol. 16 (eds Hechet, M. K., Wallace, B. & Prance, G. T.) (Plenum, New York, 1983).
22. von Kiedrowski, G. in *40 Jahre Fonds der Chemischen Industrie 1950-1990*, 197-218 (Verband der Chemischen Industrie, Frankfurt, 1990).
23. von Kiedrowski, G. *et al. Nachr. Chem. Tech. Lab.* (in the press).
24. Doudna, J. A., Couture, S. & Szostak, J. W. *Science* **251**, 1605-1610 (1991).
25. Terfort, A. & von Kiedrowski, G. *Angew. Chem. int. Ed. Engl.* **31** (in the press).
26. Inoue, T. & Orgel, L. E. *J. Am. chem. Soc.* **103**, 7666-7667 (1981).
27. Inoue, T. & Orgel, L. E. *Science* **219**, 859-862 (1983).
28. Wu, T. & Orgel, L. E. *J. Am. chem. Soc.* **114** (in the press).
29. Acevedo, O. L. & Orgel, L. E. *J. molec. Biol.* **197**, 187-193 (1987).
30. Kang, CH., Zhang, X., Ratliff, R., Moyzis, R. & Rich, A. *Nature* **356**, 126-131 (1992).
31. Challa, G. & Tan, Y. Y. *Pure appl. Chem.* **53**, 627-641 (1981).
32. van de Grampel, H. T., Tuin, G., Tan, Y. Y. & Challa, G. *Macromolecules* **25**, 1049-1056 (1992).
33. Ballard, D. G. H. & Bamford, C. H. *Proc. R. Soc.* **A236**, 384 (1956).
34. Bamford, C. H. in *Reactions on Polymers*, *Proc. NATO Adv. Study Inst.* (ed. Moore, J. A.) 54-60 (Reidel, Dordrecht, 1973).
35. Shapiro, R. *Origins. A Skeptic's Guide to the Creation of Life on Earth* (Summit, New York, 1986).
36. Müller, D. *et al. Helv. chim. Acta* **73**, 1410-1468 (1990).
37. Joyce, G. F. *et al. Nature* **310**, 602-604 (1984).
38. Hill Jr, A. R., Nord, L. D., Orgel, L. E. & Robins, R. K. *J. molec. Evol.* **28**, 170-171 (1989).
39. Visscher, J. & Schwartz, A. W. *J. molec. Evol.* **26**, 291-293 (1987).
40. Harada, K. & Orgel, L. E. *Origins Life* **20**, 151-160 (1990).
41. Harada, K. & Orgel, L. E. *J. molec. Evol.* **32**, 358-359 (1991).
42. Brack, A., Ehler, K. W. & Orgel, L. E. *J. molec. Evol.* **8**, 307-310 (1976).
43. Joyce, G. F. & Orgel, L. E. *J. molec. Biol.* **188**, 433-441 (1986).
44. Joyce, G. F., Inoue, T. & Orgel, L. E. *J. molec. Biol.* **176**, 279-306 (1984).
45. Eigen, M. *Naturwissenschaften* **58**, 465-532 (1971).
46. Kaufmann, S. A. *J. theor. Biol.* **119**, 1 (1986).
47. Wächtershäuser, G. *Microbiol. Rev.* **52**, 452-484 (1988).
48. De Duve, C. *Blueprint for a Cell: The Nature and Origin of Life* (Patterson, Burlington, North Carolina, 1991).
49. von Kiedrowski, G. *et al. Origins Life* **22** (in the press).
50. Cairns-Smith, A. G. *Genetic Takeover and the Mineral Origins of Life* (Cambridge Univ. Press, 1982).
51. Cairns-Smith, A. G. & Davies, C. J. in *Encyclopaedia of Ignorance* (eds Duncan, R. & Weston-Smith, M.) 391-403 (Pergamon, Oxford, 1977).
52. Weber, A. L. *Origins Life* **17**, 107-119 (1987).
53. Orgel, L. E. *J. molec. Biol.* **38**, 381-393 (1968).
54. Weber, A. L. *Origins Life* **19**, 179-186 (1989).
55. Nelsestuen, G. L. *J. molec. Evol.* **15**, 59-72 (1980).
56. Schwartz, A. W. & Orgel, L. E. *Science* **228**, 585-587 (1985).
57. Egholm, M., Buchardt, O., Nielsen, P. E. & Berg, R. H. *J. Am. chem. Soc.* **114**, 1895-1897 (1992).
58. Bosscher, F., Ten Brinker, G. & Challa, G. *Macromolecules* **15**, 1442-1444 (1982).

ACKNOWLEDGEMENTS. I thank A. Weber for drawing my attention to literature on template-directed vinyl polymerizations, A. Hill for drawing the figures and S. Bailey for manuscript preparation. This work was supported by NASA and NSCORT NASA. Figure 6 is reprinted with permission from ref. 58.

Atomic structure and chemistry of human serum albumin

Xiao Min He & Daniel C. Carter*

The Space Science Laboratory, ES76 Biophysics, Marshall Space Flight Center, Huntsville, Alabama 35812, USA

The three-dimensional structure of human serum albumin has been determined crystallographically to a resolution of 2.8 Å. It comprises three homologous domains that assemble to form a heart-shaped molecule. Each domain is a product of two subdomains that possess common structural motifs. The principal regions of ligand binding to human serum albumin are located in hydrophobic cavities in subdomains IIA and IIIA, which exhibit similar chemistry. The structure explains numerous physical phenomena and should provide insight into future pharmacokinetic and genetically engineered therapeutic applications of serum albumin.

THE serum albumins belong to a multigene family of proteins that includes α -fetoprotein (AFP) and human group-specific component (Gc) or vitamin D-binding protein. They are relatively large multi-domain proteins which, as the major soluble protein constituents of the circulatory system, have many physiological functions. The albumins contribute significantly to colloid osmotic blood pressure and aid in the transport, distribution and metabolism of many endogenous and exogenous ligands. These ligands represent a spectrum of chemically diverse molecules, including fatty acids, amino acids (notably tryptophan and cysteine), steroids, metals such as calcium, copper and zinc, and numerous pharmaceuticals. They are implicated

in the facilitated transfer of many ligands across organ-circulatory interfaces such as in the liver, intestine, kidney and brain¹, and evidence suggests the existence of an albumin cell surface receptor². In addition to blood plasma, serum albumins are also found in tissues and bodily secretions throughout the body; the extravascular protein comprises 60% of the total albumin. Unlike AFP or Gc, albumins are not glycosylated and play no role in immunosuppression. Human serum albumin (HSA), a protein of M_r 65K, consists of 585 amino acids. Its amino-acid sequence contains a total of 17 disulphide bridges, one free thiol (Cys 34), and a single tryptophan (Trp 214). The disulphides are positioned in a repeating series of nine loop-link-loop structures centred around eight sequential Cys-Cys pairs. A total of 61% of the amino-acid sequences are conserved among the known

* To whom correspondence should be addressed.

sequences of bovine³, rat⁴ and human serum albumins⁵. More recently, several additional sequences have been determined including sheep⁶, frog⁷, salmon⁸, mouse⁹, pig¹⁰ and sea lamprey¹¹. Sequences are also known for α -AFP¹²⁻¹⁴, and Gc proteins¹⁵⁻¹⁷ of human, rat and mouse origin. Most of these proteins share high sequence homology and all of them share the characteristic repeating series of disulphide bridges. All members of the albumin multigene family for which sequences have been determined have internal sequence homology (from two- to sevenfold), suggesting the proteins evolved from a common ancestral protein of about 190 amino acids¹⁸. This internal structural homology has been previously verified by low-resolution crystallographic studies with HSA^{19,20}. (For reviews see refs 21-24.)

In this paper, we report the complete atomic model of two crystal forms of human serum albumin as determined crystallographically at 3.1 Å for the wild-type HSA and 2.8 Å for a recombinant form expressed in yeast (rHSA)²⁵. Additionally, the characteristic binding locations and chemistry for a selection of representative biological and pharmaceutical ligands are presented.

Structure determination

Crystals of HSA used in this study were obtained from solutions of defatted albumin in polyethylene glycol (PEG) with an M_r of 400 at neutral pH as previously described¹⁹. These crystals are unusual because: (1) they grow in the rarely observed space group $P4_22_1$; (2) they have continuous solvent channels centred on the 2-fold axes parallel to the crystallographic c -axis which have a cross-section of ~ 90 Å; and (3) the crystals have a very high solvent content of 78%. The crystals are weak scatterers of X-rays and although selected crystals may show diffraction to d -spacings less than 3.0 Å, crystallization and X-ray diffraction experiments indicate a useful limit of 3.1 Å. Crystals of recombinant human serum albumin (rHSA) were obtained from solutions of PEG at neutral pH, belong to the space group $P2_1$, and diffract X-rays to d -spacings of < 2.4 Å. The structures of HSA and rHSA were determined by isomorphous and molecular replacement methods, respectively, as outlined in Table 1 and Fig. 1.

Molecular configuration

The topology of HSA is created by a repeating series of six helical subdomains²⁰. Consequently, the serum albumin multigene family represents, taxonomically, a distinct structural class of proteins. These six subdomains assemble to form a heart-shaped molecule identical in size and shape to the low-resolution electron micrographs of the 39% homologous human and bovine α -AFP²⁶ (Fig. 2). The shape is markedly asymmetric, and can be approximated to a solid equilateral triangle with sides of ~ 80 Å and average depth of ~ 30 Å. Consistent with the internal amino-acid sequence homology, there are three structurally homologous domains that repeat in the molecule (denoted I, II and III). Each domain is formed by two smaller subdomains, A and B, as previously described at low resolution^{19,20}. Altogether, roughly 67% of HSA is helical, with the remainder in turns and extended polypeptide. There are 10 principal helices in each domain (h1-h10) (Fig. 3a). The topology of a typical domain is further illustrated in Fig. 3b. Subdomain A and B share a common motif that includes h1, h2, h3 and h4 for subdomain A, and h7, h8, h9 and h10 for subdomain B. One exception is that the disulphide bridge connecting h1 and h3 does not exist in subdomain IA. The r.m.s. distances between the corresponding α -carbons of the common motifs of subdomains A and B were determined by least-squares fitting to be 2.47 Å, 2.53 Å and 2.60 Å (57 atom pairs) for domains I, II and III, respectively. In addition to this common motif, subdomain A is supplemented by two additional short antiparallel helices, h5 and h6, which are tied together by a pair of disulphide bridges forming a smaller disulphide double loop (loops 2, 5 and 8,

described in Fig. 3). Together, these additions form a virtually continuous helical globin-like fold for the A subdomains which is extensively crosslinked by a total of four interhelical disulphide bridges. The B subdomains supplement the helical motif with an N-terminal portion of extended polypeptide, which creates a folding topology closely resembling a simple up-down helical bundle. The intradomain connections between subdomains IA-IB, IIA-IIB and IIIA-IIIB consist of extended polypeptide from residues Lys 106 to Glu 119, Glu 292 to Val 315, and Glu 492 to Ala 511 in domains I, II and III, respectively. In contrast, the interdomain connections represent a helical continuation from the C-terminal portion of IB and IIB helices to the N-terminal helices of IIA and IIIA (helices: h10(I)-h1(II) and h10(II)-h1(III)), respectively. In the case of the IB-IIA connection this results in the largest helix in the structure (h10(I)-h1(II)) which contains 31 residues.

The disulphide pairings in the primary structure of HSA occur as predicted by Brown⁶. In the three-dimensional structure, the 17 cystines form disulphide linkages which occur primarily between α -helices, often distorting the local helical conformation. It is notable that the occurrence of interhelical disulphides has been rarely observed in protein structure based on a cursory survey of the Brookhaven Protein Data Bank. As predicted by Raman spectroscopic studies²⁷, the conformations of the thioether linkages are predominantly gauche-gauche-gauche, and typical $C\beta_1-S_1-S_2-C\beta_2$ torsion angles cluster around ± 80 degrees.

Ligand binding

It became evident from early heavy-atom derivative screens and from preliminary binding studies that the principal binding regions on HSA were located in subdomains IIA and IIIA. As a result, the relative binding locations have now been determined crystallographically for several ligands at low resolutions (Table 2). The binding cavity in IIIA is the most active and accommodating on HSA, many ligands were found to bind preferentially there, for example, digitoxin, ibuprofen and tryptophan. Aspirin and iodinated aspirin analogues show nearly equal distributions between binding sites located in IIA and IIIA. Warfarin occupies a single site in IIA. These observations agree with the predicted locations based on competitive inhibition and spectroscopic studies. Residues Trp 214, Lys 199 and Tyr 411 have been implicated in the binding process by several studies²⁸⁻³², and each is located strategically in the IIA or IIIA hydrophobic pockets.

For brevity, only the binding chemistry of ligand 2,3,5-triiodobenzoic acid (TIB) is discussed in detail. TIB, as with many small aromatic carboxylic acids, is bound more or less equally in both IIA and IIIA. Additionally, this ligand serves to illustrate the differences and similarities of the protein-ligand interactions between the two subdomains. In subdomain IIA (Fig. 4a) TIB is bound in a hydrophobic crevice in the cavity. The aromatic ring forms hydrophobic interactions with residues Leu 219, Phe 223, Leu 234, Leu 238, Leu 260, Ala 261, Ile 264, Ile 290, Ala 291 and the hydrocarbon chain of Glu 292. The carboxylate group interacts with Arg 257, Arg 222 and Lys 199, and is within ~ 4.0 Å of His 242. The distribution of hydrophobic and hydrophilic residues in the binding crevice is distinctly asymmetric (Fig. 4a). In the IIIA binding pocket (Fig. 4b) the hydrophobic portion of the aromatic ring of TIB is packed against Pro 384, Leu 387, Ile 388, Phe 395, Leu 407, Leu 430, Val 433, Ala 449, Leu 453 and the hydrocarbon chains of Arg 485 and Glu 450. The carboxylate interacts primarily with Arg 410, and is within 4.0 Å of the oxygen of Tyr 411.

Although the chemistry of ligand binding of IIIA is analogous to that of IIA, there are marked differences in the location of bound ligand. In subdomain IIA, TIB is located in that part of the cavity created by the smaller disulphide double loop (h5(II) and h6(II)); whereas in IIIA, it is located much closer to h1(III). When salicylic acid derivatives or other small aromatic

TABLE 1 Heavy-atom derivatives, phasing methods and statistics

Data set		Atom	X	Y	Z	O _c	R _c	R _f	R _k	N	f _{h/e}	D
PIP	1	Pt	0.3878	0.0797	0.1498	1.07	0.59	0.22	0.16	6778	1.31	4.0
		Pt	0.0258	0.3863	0.2275	0.37						
		Pt	0.2298	0.4331	0.3037	0.77						
		Pt	0.2940	0.0673	0.0803	0.79						
	2					0.53	0.17	0.11	3645	1.72	4.0	
	3					0.59	0.18	0.11	2641	1.66	5.6	
HgI ₂	1	Hg	0.1568	0.3595	0.0660	0.38	0.62	0.21	0.16	4565	1.25	4.0
		Hg	0.1857	0.2778	0.0615	1.11						
		Hg	0.0873	0.3761	0.1397	0.78						
		Hg	0.2443	0.2172	0.1718	0.30						
	2					0.65	0.20	0.13	4334	1.22	4.0	
	3					0.73	0.18	0.12	1557	1.15	5.0	
IS	1	I	0.0797	0.3778	0.1378	0.73	0.64	0.18	0.12	3876	1.47	4.0
		I	0.1811	0.2695	0.0751	0.71						
	2					0.66	0.20	0.15	4043	1.08	4.0	
	3					0.69	0.20	0.15	4108	1.17	4.0	
	4					0.66	0.19	0.14	6334	1.27	4.0	
	5					0.68	0.18	0.14	6715	1.09	4.0	
DIS	1	I	0.1820	0.2716	0.0720	0.77	0.67	0.20	0.14	4734	1.29	4.0
		I	0.0819	0.3756	0.1289	0.91						
TAM	1	Hg	0.0556	0.4554	0.2112	0.64	0.69	0.13	0.11	6334	1.27	4.0
		Hg	0.2088	0.4376	0.4413	0.57						
	2					0.74	0.17	0.12	4277	0.52	4.0	
	3					0.73	0.17	0.10	6277	0.66	4.0	
TIB	1	I	0.1827	0.2727	0.0761	1.32	0.61	0.22	0.16	5431	1.36	4.0
		I	0.0830	0.3753	0.1248	1.37						
HgCl ₂	1	Hg	0.0578	0.4540	0.2217	1.04	0.49	0.26	0.14	2201	1.89	6.8
		Hg	0.3429	0.1026	0.3494	0.40						
		Hg	0.1547	0.4589	0.2035	0.49						
RuCl ₂	1	Ru	0.0784	0.3732	0.1334	0.18	0.75	0.17	0.15	2490	0.47	6.0
		Ru	0.1003	0.3125	0.2495	0.36						
		Ru	0.1916	0.2999	0.3024	0.27						
K ₂ PtCl ₆		Pt	0.1039	0.3686	0.1315	0.57	0.74	0.11	0.10	5883	0.58	4.0
		Pt	0.1535	0.2959	0.3892	0.10						
MER	1	Hg	0.2499	0.2159	0.1443	0.67	0.69	0.11	0.12	7778	0.91	4.0
		Hg	0.1665	0.3367	0.3718	0.73						
	2					0.68	0.11	0.12	3200	0.88	4.0	

The crystal structure of HSA (space group $P4_21_2$; unit cell constants: $a=b=186.5(5)$ Å, $c=81.0(5)$ Å) was phased to 4.0 Å resolution using multiple isomorphous replacement (MIR) data collected from 12 heavy-atom derivatives tabulated above. To prepare heavy-atom-HSA complexes, crystals of HSA were placed in a stabilizing solution of 45% PEG 400 (pH 6.8–7.0) which contained about 0.5 to 5 mM of the compound of interest and allowed to stand for 24 to 72 h. X-ray diffraction data were collected on a Siemens multiwire area detector using a rotating anode source operating at 40 kV and from 70 to 90 mA. Data were reduced using the Xengen program package operating on a microVAX III⁴². Diffraction data used for the isomorphous replacement studies, and initial model construction and refinement were collected from three crystals and merged to give $R_{\text{merge}}=11.6\%$ based on intensity for 225,950 observations of 24,249 unique reflections (17,575 unique reflections $>1\sigma F$). Native data used in the final refinement were collected from a selected crystal grown in microgravity (STS-42), $R_{\text{merge}}=9.25\%$ based on 49,113 observations of 27,096 unique reflections (21,622 unique reflections $>1\sigma F$). These data were used in the current refinement of HSA. Difference maps ($F_{\text{PH,obs}} - F_{\text{P,calc}}$) were calculated to identify the binding loci using phases obtained from MIR and solvent flattening as previously described⁴⁹. Because of problems with the radiation stability of the heavy-atom complexes, several partial data sets were required to complete the data on a number of derivatives. Each data set was refined separately: first by least-squares against the centric data, then by an iterative series of lack-of-closure refinements⁴³ (only one set of refined coordinates are shown in the table for each derivative). This resulted in the combination of 21 individual data sets that were selected from >100 low-to-medium resolution data sets collected with a multiwire area detector. These data produced a mean figure of merit = 66.8 for 9,253 paired reflections greater than 5σ excluding anomalous data and 0.71 including anomalous data from eight of the derivatives. In many cases it was not possible to distinguish the individual heavy atoms in some complexes, such as TAM, PIP; consequently, the complexes were treated as individual heavy atoms (usually with larger B-factors) in the refinement. Solvent flattening was used to improve further the MIR phases⁴⁴. From the resulting electron density, the topology of the structure could be unambiguously traced, and the positions of all 17 disulphide bridges were clearly recognizable. At this time, a polyaniline model (including the 17 disulphides) of HSA was constructed from the electron density using computer graphics⁴⁵, and refined by simulated annealing (SA)⁴⁶ to $R=0.29$ at 3.5 Å. After an extensive series of unbiased omit maps in which 40 residues were deleted at one time, the complete amino-acid sequence, with the exception of residues 1–3, was incorporated and refined to 3.2 Å against the native data (Fig. 1a). The current crystallographic R -factor is 24.5% for 15,594 reflections from 6.0 to 3.1 Å with $F > 2\sigma$. Current r.m.s. deviations of the model from ideality are 0.028 Å for bond lengths and 5.3 degrees for bond angles. Crystals of rHSA belong to the monoclinic space group $P2_1$ with cell constants: $a=58.9$ Å, $b=38.3$ Å, $c=60.7$ Å and $\beta=101.9$ degrees. In contrast to the tetragonal crystal form above, these crystals have solvent contents of ~33%. Diffraction data were collected from a single crystal and merged to give $R_{\text{merge}}=8.0\%$ based on intensity for 14,772 unique data of 51,680 observations to 2.27 Å (97% complete to 2.8 Å). The atomic model of tetragonal HSA determined as above, was used to solve the structure of rHSA by the method of molecular replacement⁴⁷. Rotation function^{48,49} and translation studies⁵⁰ using the program package MERLOT⁵¹ yielded single high-contrast solutions. The preliminary molecular replacement model produced an R -factor of 44% and a correlation coefficient (cc) of 58% without refinement for 4,979 reflections with structure factors greater than 3σ in the resolution range of 8.0 Å to 4.0 Å. The model was subjected to rigid-body refinement (in six segments) and then was refined by SA to give an R -factor of 24.5% for 11,885 reflections with $F > 1\sigma$ in the resolution range of 10.0 Å–3.0 Å. This model was refit in an analogous manner as described for HSA using omit maps (Fig. 1b) and subjected to further SA refinement to give $R=23.8\%$ for 10,672 reflections from 6.0 to 2.8 Å with $F > 2\sigma$. R.m.s. deviations of the model from ideality were 0.023 Å for bond lengths and 4.9 degrees for bond angles.

O_c, Relative occupancy.R_c, Centric r -factor (Cullis), $\sum ||F_{\text{PH,obs}} \pm F_{\text{P,obs}} - F_{\text{H,calc}}| / \sum |F_{\text{PH,obs}} \pm F_{\text{P,obs}}|$.R_f, $\sum |F_{\text{PH}} - F_{\text{P}}| / \sum F_{\text{P}}$.R_k, Kraut R -factor $\sum ||F_{\text{PH,obs}}| - |F_{\text{H,calc}}|| / \sum |F_{\text{PH,obs}}|$.N, Number of unique paired reflections $> 5\sigma$.f_{h/e}, Phasing power, $(\sum |F_{\text{H,calc}}|^2 / \sum (|F_{\text{PH,obs}}| - |F_{\text{H,calc}}|)^2)^{1/2}$.

MER, Mersalyl acid, 2-(N-(3-hydroxymercuri-2-methoxypropyl)carbanoyl)phenoxyacetic acid.

cc, $(\langle F_o F_c \rangle - \langle F_o \rangle \langle F_c \rangle) / ((\langle F_o^2 \rangle - \langle F_o \rangle^2)^{1/2} (\langle F_c^2 \rangle - \langle F_c \rangle^2)^{1/2})$.

D, Resolution of data in Å.

PIP, Di-μ-iodobis(ethylenediamine)-di-platinum(II) nitrate.

IS, 5-Iodosalicylic acid.

DIS, 3,5-Diiodosalicylic acid.

TAM, Tetakis(acetoxymethyl)mercuric methane.

TIB, 2,3,5-Triiodobenzoic acid.

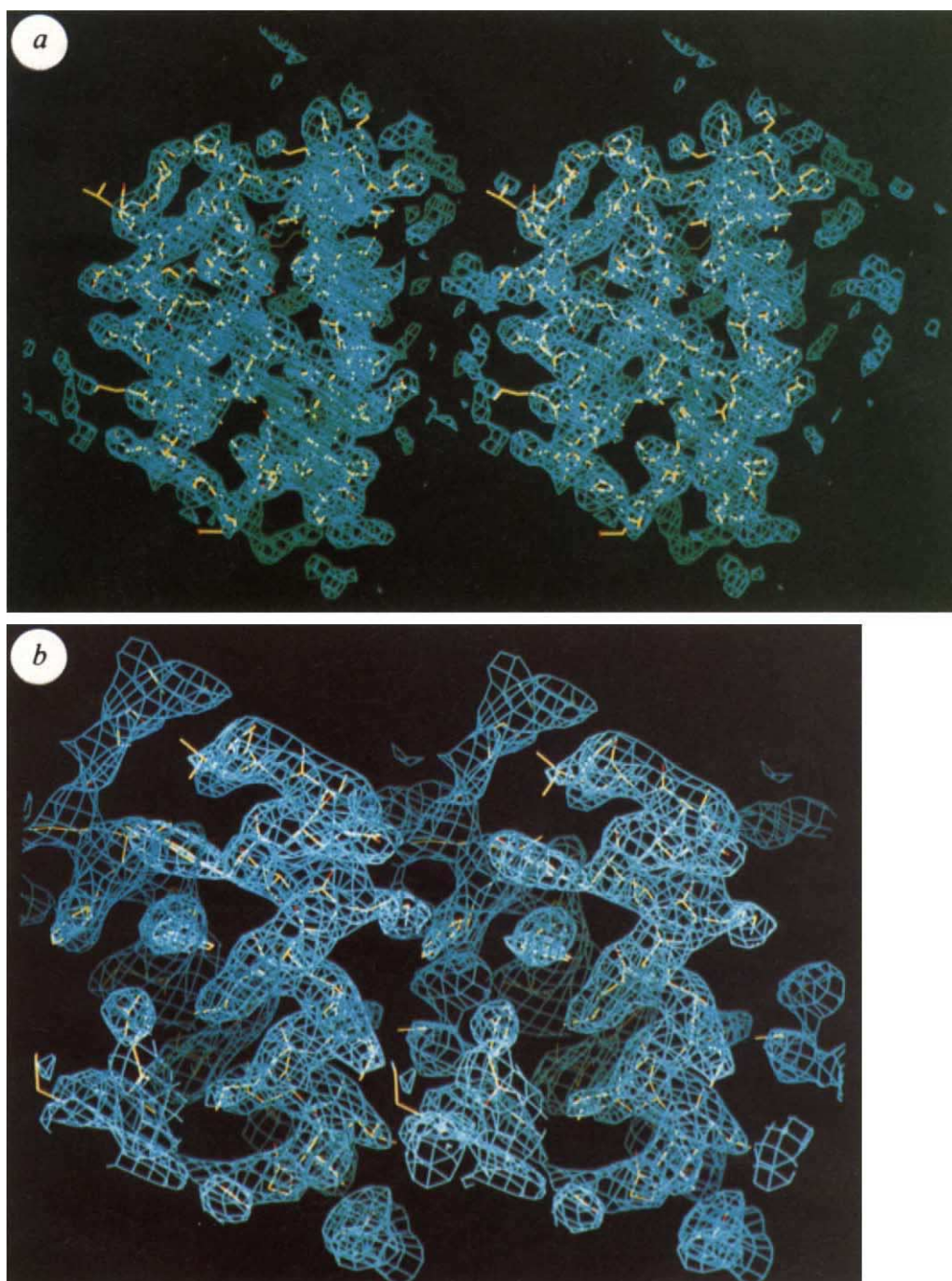
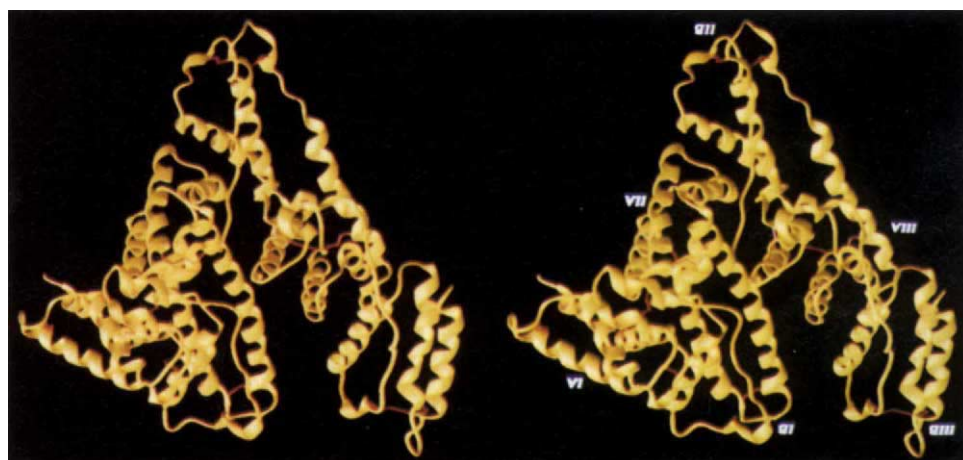


FIG. 1 *a*, Omit map produced from the deletion of a helical motif (residues 196–272) from the model ($\frac{1}{2}$ of the structure). The difference map is shown at 3.2 Å with a 1.8σ contour level (σ is the square root of the map variance). *b*, $2F_o - F_c$ electron density of the helical region from residues 344–355 of rHSA at 2.8 Å.

FIG. 2 Stereo view of human serum albumin illustrating the overall topology and secondary structure. The positions of the 17 disulphides and the side chain of Cys 34 are shown in red. Structurally, HSA consists of 28 helices which range in size from 5 to 31 amino acids in length, and can be grouped into 10 principal helices within each domain. This figure was produced with the program Ribbons⁵².



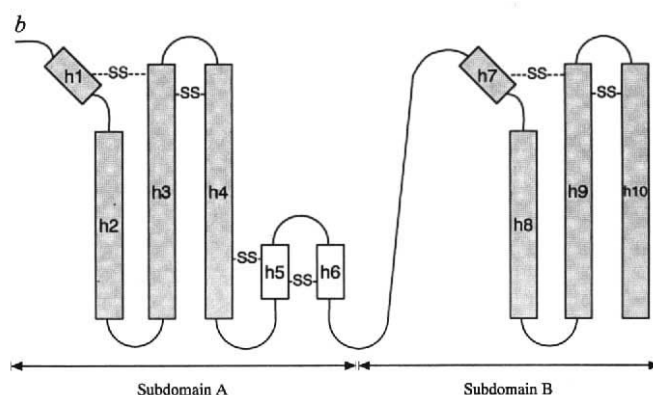
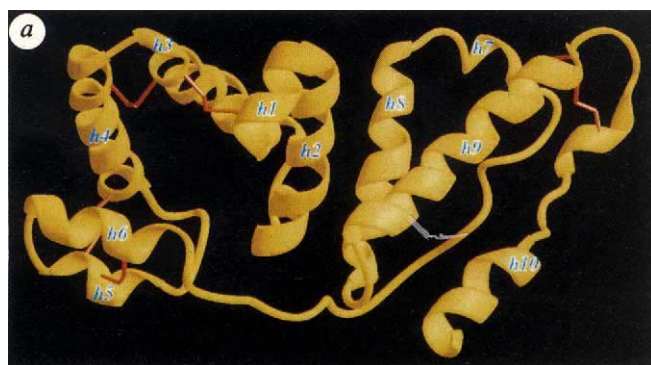


FIG. 3 *a*, Ribbon diagram of domain II. Disulphides are shown in red. The 10 principal helices in each domain are labelled as h1 to h10. *b*, Topological illustration of a typical domain in HSA. The shaded areas represent the common structural motif shared by all six subdomains. This motif is characterized by a short helix (or a short stretch of helix for subdomains IIA and IIIA) followed by three consecutive antiparallel long helices tied together by a pair of disulphide bridges. The principal portion of each subdomain is formed by one of the six larger disulphide double loops in the amino acid sequence, often referred to in the literature as loops 1, 3, 4, 6, 7 and 9 or

1A, 1C, 2A, 2C, 3A and 3C²². Three of these loops, 1, 4 and 7 are each supplemented by an additional, but smaller disulphide double loop (2, 5 and 8, respectively). Consequently the combinations of loops 1 and 2, 4 and 5, 7 and 8 represent the three A subdomains (IA, IIA and IIIA) and loops 3, 6 and 9 the B subdomains (IB, IIB and IIIB). The two largest helices in the structure occur at the interdomain connections (IB–IIA and IIB–IIIA) and represent the merged C-terminal and N-terminal helices of both domains. Thus the total number of helices is 28 rather than 30.

FIG. 4 *a*, Stereo view of subdomain IIA illustrating the chemistry of ligand binding. α -Carbons are shown in green, hydrophobic residues in yellow and hydrophilic residues in red (see text). The difference density ($F_d - F_n$) _{α} (where F_d represents the measured structure factors for the TIB complex, F_n is the native data, and the phases, α , are from the model) for TIB is shown contoured at 4σ . The N and C termini are located in the lower and upper left-hand sides of the figure, respectively. Lys 199 and His 242 are located at the bottom of the figure. *b*, Stereo view of domain IIIA illustrating the chemistry of TIB binding. Colour scheme and contour of electron density are the same as in *a*. The N and C termini are located in the lower and upper right-hand side of the figure, respectively. Arg 410 and Tyr 411 are shown in red at the upper left side of the TIB binding site. The binding pockets are accessed through openings of ~ 10 Å in diameter between helices h1 and h2. Except for three to four residues in the cavities and those surrounding the opening, the pockets are lined with hydrophobic residues. For IIA, these hydrophobic residues are Leu 203, Phe 211, Trp 214, Ala 215, Leu 219, Phe 223, Leu 234, Val 235, Leu 238, Val 241, Leu 260, Ala 261, Ile 264, Ile 271, Leu 275 and Ile 290. The principal hydrophobic residues in IIIA are Pro 384, Leu 387, Ile 388, Phe 395, Leu 407, Val 415, Val 418, Leu 423, Val 426, Leu 430, Val 433, Leu 453, Val 456, Leu 457, Leu 460, Val 473 and Phe 488. These hydrophobic residues form a focal point around Lys 199 and His 242 in IIA and Tyr 411 and Arg 410 in IIIA. These figures were produced using TOM/FRODO⁵³.

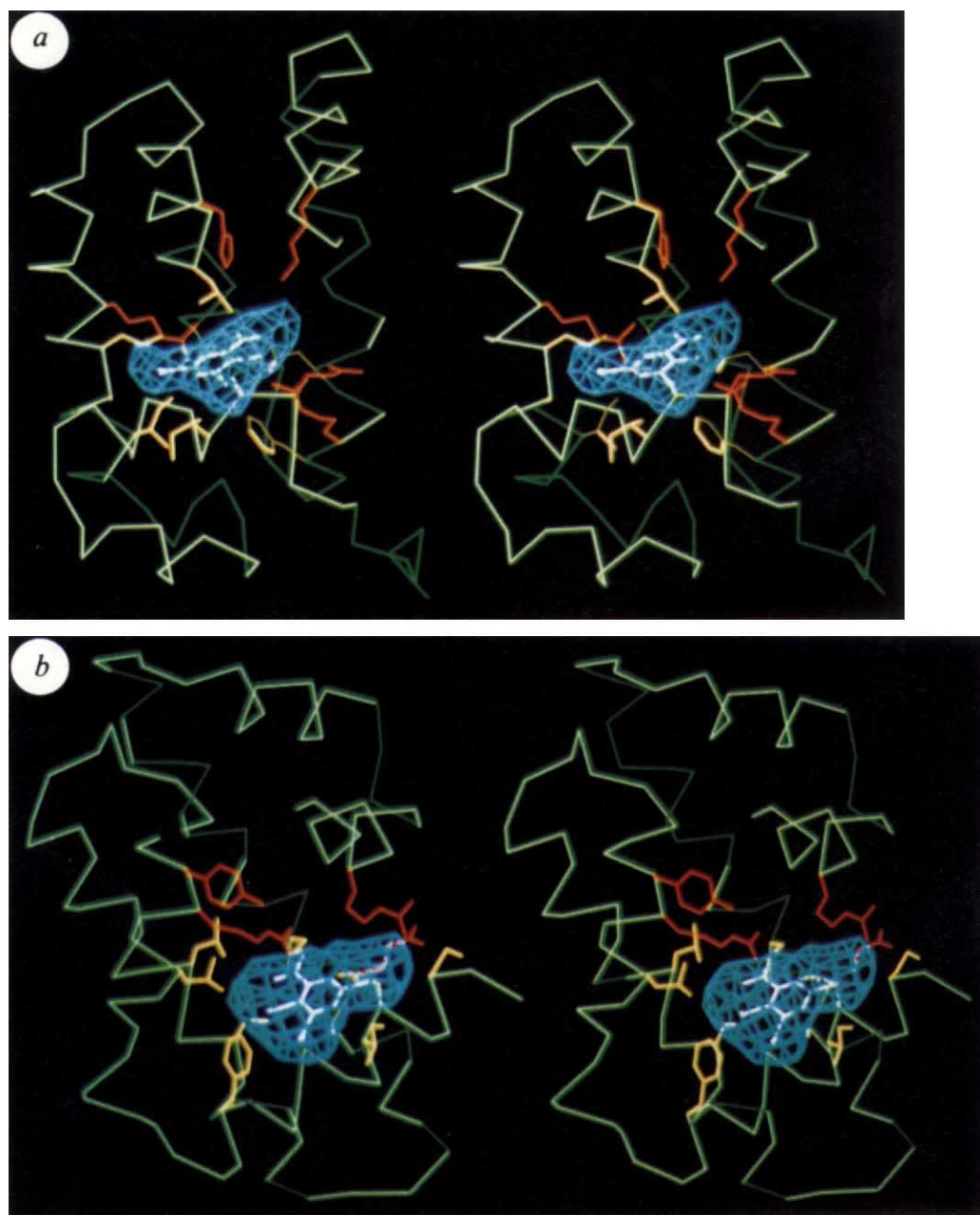


TABLE 2 Ligand binding locations to HSA

Ligand	D	N	R _t	Observed location
Aspirin	4.0	7362	0.11	IIA IIIA
Warfarin	5.0	2555	0.167	IIA
Diazepam	6.8	2075	0.118	IIIA
Digitoxin	5.0	3751	0.137	IIIA
Clofibrate	6.0	2175	0.138	IIIA
Ibuprofen	6.0	2402	0.215	IIIA
AZT	4.0	7548	0.124	IIIA
IS	4.0	6334	0.19	IIA IIIA
DIS	4.0	4734	0.20	IIA IIIA
TIB	4.0	5431	0.12	IIA IIIA

Ligand-HSA complexes and X-ray diffraction data were obtained in a manner as previously described in Table 1. The observed locations refer to the primary binding sites.

D, Resolution or d-spacing in Å.

N, Number of paired unique reflections with $F > 5\sigma$.

R_t, $\Sigma|F_{PH} - F_P|/\Sigma F_P$.

AZT, 3'-Azido-3'-deoxythymidine.

IS, 5-Iodosalicylic acid.

DIS, 3,5-Diiodosalicylic acid.

TIB, 2,3,5-Triiodobenzoic acid.

compounds are compared with TIB, there are differences in the orientation of the aromatic rings and in the details of the interactions of the carboxyl groups depending on the substitutions in the aromatic rings. But in general we found that all of the compounds listed in Table 2 were bound in the same locations within the respective subdomains with similar chemistry.

Discussion

The structures of HSA and rHSA have been determined crystallographically. The higher diffracting monoclinic rHSA crystal is virtually identical with the tetragonal HSA structure. Recently, crystals of the monoclinic form of comparable quality have been produced with HSA, which indicates that the observed structural differences may be attributed to crystal packing forces. These crystals should provide an avenue for future high-resolution refinement of the structure and its complexes. Subsequent discussion, however, will be limited to the crystallographic structure of tetragonal HSA, which explains numerous phenomena regarding the structure and chemistry of serum albumin. For example, the crystal structure explains why ligands bound to domain III affect conformational changes, as well as binding affinities, in domain II, because the binding subdomains share a common interface. Trp 214, conserved in mammalian albumins, plays an important structural role in the formation of the IIA binding site by limiting the solvent accessibility, and it participates in an additional hydrophobic packing interaction between the IIA and IIIA interface. Both Lys 199, which can be acetylated by aspirin^{33,54}, and Tyr 411, which can be esterified by *p*-nitrophenyl acetate³⁰⁻³², are located in the primary binding pockets close to bound ligands. It is now understandable why proteolytic cleavage of HSA would produce two halves of the molecule that can reassociate in solution, thereby restoring the binding properties of the intact albumin³⁴. In addition, the amino-acid sequences of serum albumin recognized by the proteases, trypsin and pepsin, lie on the surface of the molecule in flexible solvent accessible inter-subdomain connections. Similarly, a variety of rare, naturally occurring single-site point mutations of HSA identified by anomalous electrophoretic migration are located on the surface of the molecule and exposed to solvent³⁵⁻³⁷.

Ligand binding to albumin has been studied for over 50 years. Generally, serum albumin has a greater affinity for small, negatively charged hydrophobic molecules. In this regard it is notable

that serum albumin is the principal carrier of fatty acids in the blood which are otherwise insoluble. On the basis of studies with proteolytic fragments of both human and bovine serum albumin³⁸, the principal binding site for long-chain fatty-acids has been shown to occupy the third domain. In a similar manner, the high-affinity site for bilirubin has been isolated in domain II³⁹. Thus, one can infer that the primary long-chain fatty-acid and bilirubin-binding sites reside within IIIA and IIA, respectively. Cys 34, the single free thiol of HSA, is the location of bound cysteine and glutathione²², and also forms complexes with various mercurial and gold compounds, such as the anti-arthritis auranofin⁴⁰. In the structure, Cys 34 is partially protected from the solvent, located in a turn between helices h2(I) and h3(I). The strong binding of other metals, such as Cu(II) and Ni(II), occurs only in albumins with a histidine at residue 3⁴¹. Unfortunately, in both crystal structures, residues 1-3 were not observed in the electron density, which suggests that this is a flexible region of the molecule.

The absence of observed ligand binding by homologous subdomain IA can be explained by the structural differences between this subdomain and subdomains IIA and IIIA. In IA there is a region of extended polypeptide after Cys 62, which allows helix h4(I) to adjust its packing interaction with h3(I) (which is near zero degrees in equivalent helix pairs, such as h3(II) and h4(II)), thereby effectively eliminating the potential binding pocket.

Serum albumin has long been used as a model protein and served numerous applications in both industrial processes and academic research areas. Current advances in recombinant technologies, coupled with a more complete understanding of albumin structure and function, should provide for a greater abundance of future applications. Accordingly, with the wealth of drug and ligand binding data, researchers should now be in a position to understand and predict ligand displacement interactions for a variety of endogenous and exogenous ligands. In this regard, IS, DIS and TIB should prove to be valuable tools in future ligand/drug displacement studies. Moreover, the structure should provide a basis for modifications of the protein to carry therapeutic or diagnostic agents, and create smaller proteins with enhanced binding activities for a variety of applications. □

Received 31 March; accepted 11 June 1992.

- Pardridge, W. M. *Am. J. Physiol.* **252**, 157-164 (1987).
- Schnitzer, J. E., Carley, W. W. & Palade, G. E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6773-6777 (1988).
- Brown, J. R. & Shockey, P. in *Lipid-Protein Interactions* Vol. 1 (eds Jost, P. & Griffith, O. H.) 25-68 (Wiley, New York, 1982).
- Sargent, T. D., Yang, M. & Bonner, J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 243-246 (1981).
- Duglaczky, A., Law, S. W. & Dennison, O. E. *Proc. natn. Acad. Sci. U.S.A.* **79**, 71-75 (1982).
- Brown, W. M., Dziegielewska, K. M., Foreman, R. C. & Saunders, N. R. *Nucleic Acids Res.* **17**, 10495 (1989).
- Moskatis, J. E., Sargent, T. D., Smith, L. H. Jr, Pastori, R. L. & Schoenberg, D. R. *Molec. Endocr.* **3**, 464-473 (1989).
- Byrnes, L. & Gannon, F. *DNA Cell Biol.* **9**, 647-655 (1990).
- Minghetti, P. P., Law, S. W. & Duglaczky, A. *Molec. Biol. Evol.* **2**, 347-358 (1985).
- Weinstock, J. & Baldwin, G. S. *Nucleic Acids Res.* **16**, 9045 (1988).
- Gray, J. E. & Doolittle, R. F. *Protein Sci.* **1**, 289-302 (1992).
- Moringa, T., Sakai, M., Wegmann, T. G. & Tamaoki, T. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4604-4608 (1983).
- Jagodzinski, L. L., Sargent, T. D., Yang, M., Glackin, C. & Bonner, J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3521-3525 (1981).
- Gorin, M. B. & Hoffman, B. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1351-1355 (1980).
- Schoentgen, F., Metz-Boutigue, M.-H., Jolles, J., Constants, J. & Lolles, P. *Biochim. biophys. Acta* **871**, 189-198.
- Cooke, N. E. & David, E. V. *J. clin. Invest.* **76**, 2420-2424 (1985).
- Yang, F. *et al. Genomics* **7**, 509-516 (1990).
- Brown, J. R. *Fed. Proc.* **35**, 2141-2144 (1976).
- Carter, D. C. *et al. Science* **244**, 1195-1198 (1989).
- Carter, D. C. & He, X. M. *Science* **249**, 302-303 (1990).
- Putnam, F. W. *The Plasma Proteins* 2nd edn Vol. 4 (Academic, London, 1984).
- Peters, T. Jr *Adv. Protein Chem.* **37**, 161-245 (1985).
- Fehske, K. J., Muller, W. E. & Wollert, U. *Biochem. Pharmacol.* **30**, 687-692 (1981).
- Kragh-Hannsen, U. *Pharmac. Rev.* **33**, 17-53 (1981).
- Quirk, A. V. *et al. Biotechnol. appl. Biochem.* **11**, 273-287 (1989).
- Luft, A. J. & Lorscheider, F. L. *Biochemistry* **22**, 5978-5980 (1983).
- Aoki, K., Sato, K., Nagoaka, S., Kamada, M. & Hiramatsu, K. *Biochim. biophys. Acta* **328**, 323-333 (1973).
- Sudlow, G., Birkett, D. J. & Wade, D. N. *Molec. Pharmacol.* **11**, 824-832 (1975).
- Sudlow, G., Birkett, D. J. & Wade, D. N. *Molec. Pharmacol.* **12**, 1052-1061 (1977).
- Sollene, N. P. & Means, G. E. *Molec. Pharmacol.* **14**, 754-757 (1979).

31. Ozeki, Y., Kurono, Y., Yotsuyanagi, T. & Ikeda, K. *Pharmac. Bull.* **28**, 535–540 (1980).
32. Kurono, Y., Ozeki, Y., Yamada, H., Takeuchi, T. & Ikeda, K. *Chem. Pharmac. Bull.* **35**, 734–739 (1987).
33. Hagag, N., Birnbaum, E. R. & Darnall, D. W. *Biochemistry* **22**, 2420–2427 (1983).
34. Feldoff, R. C. & Ledden, D. J. *Bioch. biophys. Res. Commun.* **114**, 20–27 (1983).
35. Takahashi, N. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8001–8005 (1987).
36. Arai, K., Ishioka, N., Huss, K., Madison, J. & Putnam, F. W. *Proc. natn. Acad. Sci. U.S.A.* **86**, 434–438 (1989).
37. Galliano, M. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8721–8725 (1990).
38. King, T. P. *Arch. biochem. Biophys.* **156**, 509–520 (1973).
39. Bos, O. J. M., Remijn, J. P. M., Fischer, J. E., Witting, J. & Janssen, L. H. M. *Biochem. Pharmac.* **37**, 3905–3909 (1988).
40. Pedersen, S. M. *Biochem. Pharmac.* **35**, 2661–2666 (1987).
41. Dixon, J. W. & Sarkar, B. *J. biol. Chem.* **249**, 5872–5877 (1974).
42. Howard, A. J. et al. *J. appl. Crystallogr.* **20**, 383–387 (1987).
43. Furey, W. & Swaminathan, S. *Am. crystallogr. Assoc. Mtg. Abstr. Ser 2* **18**, 73 (1990).
44. Wang, B. C. *Meth. Enzym.* **115**, 90–112 (1985).
45. Jones, T. A. *J. appl. Crystallogr.* **20**, 383 (1987).
46. Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
47. Rossmann, M. G. & Blow, D. M. *Acta Crystallogr.* **15**, 24–31 (1962).
48. Crowther, R. H. in *The Molecular Replacement Method* (ed. Rossmann, M. G.) 173–178 (Gordon & Breach, New York, 1972).

49. Lattman, E. E. & Love, W. E. *Acta Crystallogr.* **B26**, 1854–1857 (1970).
50. Crowther, R. A. & Blow, D. M. *Acta Crystallogr.* **23**, 544–548 (1967).
51. Fitzgerald, P. M. D. *J. appl. Crystallogr.* **21**, 273–278 (1988).
52. Carson, M. J. *molec. Graphics* **5**, 103–106 (1987).
53. Cambillau, C. & Horjales, E. J. *molec. Graphics* **5**, 174 (1987).
54. Walker, J. E. *FEBS Lett.* **66**, 173–175 (1976).

ACKNOWLEDGEMENTS. We thank T. Peters Jr for comments on the manuscript, B. Chang, P. Twigg, E. Casale, Z. Krishnasami, K. Keeling, B. Barnes and J. Walraven for assistance with growing the crystals, D. Donovan, G. Roberts and D. Hecht for assistance, E. Abola for the cursory survey of disulphides in the Brookhaven Protein Data Bank, W. Furey for providing the computer program package PHASES, the crew of the First International Microgravity Laboratory (IML-1, STS-42) for their help with the microgravity crystallization experiments and R. S. Snyder for encouragement and support. Recombinant serum albumin was generously provided by Delta Biotechnologies in Nottingham, UK, diffraction data from the tetragonal form of HSA crystals were collected at several facilities and we thank R. Sweet, Y. Satow, Z. Otwinowski, D. Yang and Jan Troup for help and access to the Brookhaven National Synchrotron Light Source, the Photon Factory, Yale University, McMaster University and the Molecular Structure Corporation. This research was supported by a grant from the Office of Space Science and Applications of the National Aeronautics and Space Administration (NASA) to D.C.C.X.-M.H. is supported by NASA under a contract with the Universities Space Research Association.

LETTERS TO NATURE

A double-sided radio jet from the compact Galactic Centre annihilator 1E1740.7–2942

I. F. Mirabel*, L. F. Rodríguez†, B. Cordier*, J. Paul* & F. Lebrun*

* Service d'Astrophysique, Centre d'Etudes de Saclay, 91191 Gif-sur-Yvette, France

† Instituto de Astronomía, UNAM, Apartado Postal 70-264, 04510 México, DF, Mexico

RECENT observations^{1,2} with the γ -ray telescope SIGMA, on the GRANAT satellite, indicated that the hard X-ray source 1E1740.7–2942 may be the source of the strongest outbursts of 511-keV electron–positron annihilation radiation from the Galactic Centre region³. We have observed this source using the Very Large Array, and find that its radio structure is that of a double-sided jet emanating from a compact and variable core. The changes in flux density and spectral index of the core are correlated with variations in the hard X-ray output. The jets are symmetrical about the core, and end in edge-brightened radio lobes; they are probably a result of synchrotron emission of electrons and positrons from the compact core. Our observation suggest that 1E1740.7–2942 is a 'microquasar' stellar remnant near the Galactic Centre, which ejects positrons that travel more than a parsec before slowing and annihilating in the interstellar gas.

After the first observations with SIGMA, we initiated radio monitoring, at several wavelengths, of 1E1740.7–2942, the dominant hard X-ray source (≥ 30 keV) near the Galactic Centre⁴. The hard X-ray spectrum in its normal state resembles that of Cygnus X-1, one of the best candidates for an accreting black hole of stellar mass. During 13–14 October 1990, SIGMA detected a burst in the 300–600-keV energy range which has been

interpreted as annihilation of positrons in a hot medium of temperature ~ 40 keV; this is consistent with the temperature of the accretion disk derived from the X-ray continuum spectrum². Subsequently it was proposed^{5,6} that in addition this high-energy source injects positrons into a molecular cloud where they slow down and annihilate to produce the narrow component of the 511-keV line emission.

Our radio observations were done with the Very Large Array (VLA) at New Mexico, USA, during 10 periods of observation as part of a continuing monitoring programme coordinated with hard X-ray observations from space with GRANAT. The epoch of the VLA observations of 1E1740.7–2942, the array configuration, and the wavelengths observed are listed in Table 1. We have calibrated and analysed the VLA archive data of 1E1740.7–2942 for October 1988 and March 1989. For all observations we have used a bandwidth of 100 MHz. The absolute amplitude calibrator was 1328+307 and the phase calibrator 1748–253, whose 'bootstrapped' flux densities at 20, 6 and 3.6 cm were in the ranges of 1.13–1.17, 0.47–0.50 and 0.27–0.28 Jy, respectively.

After our first period of observation (August–September 1991) we noted a change of a factor of four in the flux density of a compact radio source seen projected inside the 12'' error circle⁷ of 1E1740.7–2942. The position of this unresolved source is: $\alpha(1950) = 17^{\text{h}} 40^{\text{m}} 43.01^{\text{s}}$, $\delta(1950) = -29^{\circ} 43' 25.5''$. By analysing source counts⁸ in radio surveys, we estimate that the probability is about 0.3% that a radio source with a flux density of 0.4 mJy is an unrelated background radio source in the 12'' error circle of the X-ray source.

In Fig. 1 we compare the time variations of this radio compact source at wavelength 6 cm with the changes observed in 1E1740.7–2942 by SIGMA⁹ at hard X-ray wavelengths (1990–92) and by HEXE on the Kvant module of the MIR station¹⁰ (20–21 March 1989). Because no significant day-to-day variations of the radio flux were detected, multiple observations during intervals smaller than a week were averaged and are represented in Fig. 1 by a single point. Variations by factors of

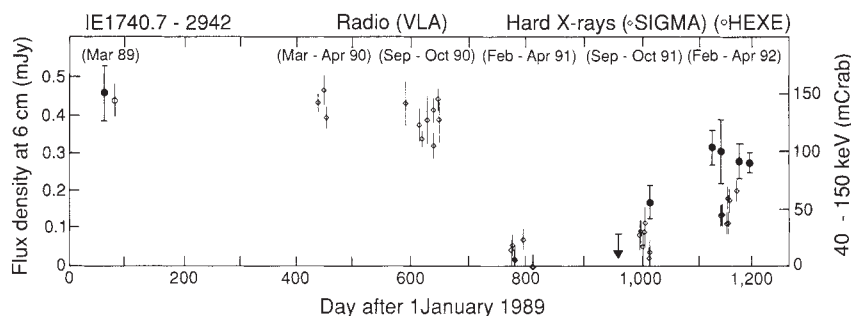


FIG. 1 VLA radio flux densities (●) at 6 cm wavelength of the radio core component, and hard X-rays measured by SIGMA (◇) and by HEXE on MIR (○) between March 1989 and April 1992.

TABLE 1 Flux densities of 1E1740.7–2942 (core source)

Epoch	VLA configuration	20 cm (mJy)	6 cm (mJy)	3.6 cm (mJy)
1988				
October 11	D/A	$<0.6 \pm 0.20$	$<0.45 \pm 0.15$	—
1989				
March 2	A/B	$<0.6 \pm 0.15$	0.47 ± 0.07	—
1991				
August 30	A	$<0.33 \pm 0.11$	$<0.09 \pm 0.03$	$<0.06 \pm 0.02$
September 3	A			
September 7	A			
October 3	A	0.31 ± 0.06	0.18 ± 0.04	$<0.09 \pm 0.03$
October 7	A/B			
1992				
January 28	C	—	0.33 ± 0.05	—
February 20	C	—	0.30 ± 0.09	—
March 21	C	—	0.27 ± 0.06	—
April 9	C	—	0.27 ± 0.03	—
April 11	C			

~ 4 were observed in the radio flux and hard X-ray photon counts. From September 1991 to February 1992, both the radio and the hard X-ray activity were enhanced. As the time variations of the radio flux density seem to be correlated with the hard X-ray intensity, we conclude that this compact source is the radio counterpart of the high-energy source 1E1740.7–2942.

The compact source was detected in several epochs at 6 cm wavelength (see Table 1), but only in October 1991 at 20 cm and not at all at 3.6 cm. Using the conventional definition for the spectral index ($S_\nu \propto \nu^\alpha$, where ν is frequency and S is flux density, from the fluxes measured on 2 March, 1989 we infer a spectral index $\alpha \geq -0.2$. From the observations of October 1991, however, we derive $\alpha \sim -0.4$. Using these results and the upper limits at 3.6 cm, we conclude that the compact source is nonthermal, with a time-variable spectral slope. It is possible that self-absorption in the core synchrotron source increases when the X-ray emission is most intense. Similar changes in spectral index have been observed in the radio counterpart of Cygnus X-1¹¹.

In February 1992 the VLA moved into the lower-resolution C configuration which provided better sensitivity for the detec-

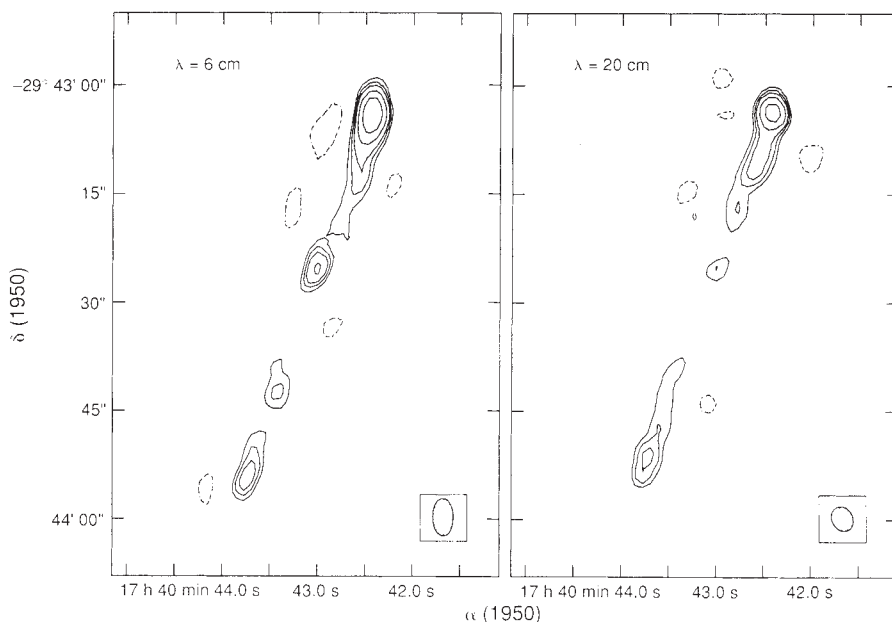
tion of weak, extended features. It then became evident that there was a double radio jet emanating from the compact source. The map at 6 cm wavelength in Fig. 2 shows that the core source is actually at the centre of a symmetrically aligned twin jet, with hotspots at both ends. The map at 20 cm wavelength shows the same general morphology with a weaker core source but stronger hotspots at the end of both the northern and southern lobes. The spectral index of the hotspots is consistent with $\alpha = -0.8 \pm 0.1$; they are likely to be enhanced synchrotron emission from charged particles accelerated anew in the strong shocks produced by the interaction of the jet with surrounding interstellar gas. In Fig. 3 we show the 20-cm map at a larger scale superposed on a molecular cloud of the Galactic Centre region projected in that direction⁶.

The integrated flux of the double radio source at 20 cm wavelength is ~ 4 mJy, about half of which comes from the northern extended hotspot. During an interval of 3 years we detected no proper motions $\geq 1.0''$ in either the central source or the northern bright spot. At a distance of 8.5 kpc, these measurements imply a total power of 3.4×10^{13} W Hz⁻¹ and proper motions $\leq 14,000$ km s⁻¹.

The morphology of the source in Fig. 2 is reminiscent of classic extragalactic radio sources, and the two-sided jet could be an unrelated background radio galaxy seen through the galactic plane. Although there is no absolute proof that this is not the case, the probability of such coincidental superposition is extremely small. The radio jets are symmetrically disposed about the compact radio source, the variation of which is correlated with that of the X-ray source. Unlike the case of Scorpius X-1, where the radio 'lobes' produced by 'unseen' jets were recently found to be unrelated background sources¹², the radio maps of 1E1740.7–2942, show, in addition to the hotspots, twin jets perfectly aligned with the core source.

1E1740.7–2942 shows some striking resemblances to Cygnus X-1. The X-ray spectra and the spectra of the core radio counterparts in both sources are similar. At the Galactic Centre, the mean X-ray and radio luminosities of Cygnus X-1 would be equal to those of 1E1740.7–2942 within a factor of two. The relatively stable radio emission from these two sources could be explained by more or less continuous conical ejection of hot plasma along twin axes perpendicular to an accretion disk¹³. This model is also consistent with our main observation of 1E1740.7–2942: the correlated variations between the radio and X-rays, the optically thick rise of the radio flux density from the central source and the two-sided jet geometry.

FIG. 2 VLA radio maps of 1E1740.7–2942 at 6 cm and 20 cm wavelength. The compact time-varying component is at the centre of a two-sided radio jet. The source integration time was ~ 12 hours at 6 cm and ~ 6 hours at 20 cm wavelength. Contours are $-3, 3, 4, 5, 7$ and 10 in units of 0.034 and 0.079 mJy per beam for the 6 cm and 20 cm maps, respectively.



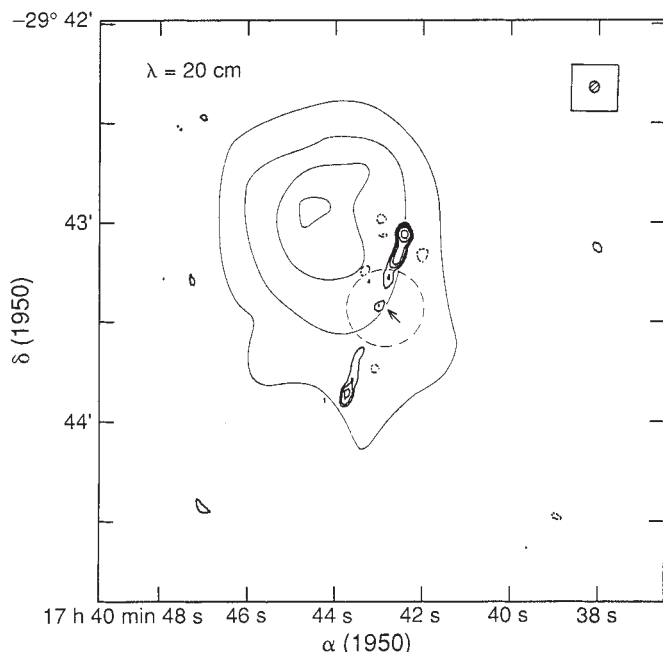


FIG. 3 VLA radio map at 20 cm wavelength of 1E1740.7-2942 (thick contours at $-3, 3, 4, 5, 7$ and 10 times 0.079 mJy per beam). The thin contours are selected from the HCO^+ map of a molecular cloud in the Galactic Centre region⁶, and the dashed line is the hard X-ray error circle⁷. The arrow marks the position of the radio core that we associate with 1E1740.7-2942.

The radio emission is likely to be synchrotron emission from relativistic charged particles ejected from the high-energy source. When the relativistic particles slow down in a dense medium, they inevitably produce high-energy γ -rays by bremsstrahlung. A high-energy γ -ray source was in fact found by COS-B¹⁴ at a position compatible with that of 1E1740.7-2942. The luminosity of this source at energies above 100 MeV is $\sim 8 \times 10^{36}$ erg s⁻¹, or 20% the total luminosity of 1E1740.7-2942.

The radio jets that we find emanating from 1E1740.7-2942 support the idea that it could be the main source of the narrow component of the 511-keV emission line from the Galactic Centre. The lack of significant redshift and the narrow width of this line indicate that the positrons annihilate in relatively cold gas far away from the high-energy source³. We propose that pairs of electrons and positrons ejected at high velocities from the compact object produce the observed synchrotron jets of radio emission. Near the Galactic Centre, their projected size of ~ 0.5 arcmin corresponds to distances of ~ 1 pc from the high-energy source. It should be noted that positrons can travel at high velocities for more than three years before they will be slowed down and annihilated in the high-density interstellar gas^{5,6} of the Galactic Centre region. $\square$

13. Hjellming, R. M. & Johnston, K. J. *Astrophys. J.* **328**, 600-609 (1988).
14. Swانبург, B. N. et al. *Astrophys. J.* **243**, L69-L73 (1981).

ACKNOWLEDGEMENTS. We thank M. Goss and B. Clark for scheduling the observations with the VLA, S. Kulkarni for permission to use his VLA archive data of 1989, A. Garcia, R. Lacambra and J. C. Yustis for help in preparing the graphical material. We appreciate the open policy of the National Radio Astronomy Observatory which is operated by Associated Universities Inc., under cooperative agreement with the U.S. NSF.

Nonlinear prediction as a way of distinguishing chaos from random fractal sequences

A. A. Tsonis* & J. B. Elsner†

* Department of Geosciences, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201, USA

† Department of Meteorology, Florida State University, Tallahassee, Florida 32306, USA

NONLINEAR forecasting has recently been shown to distinguish between deterministic chaos and uncorrelated (white) noise added to periodic signals¹, and can be used to estimate the degree of chaos in the underlying dynamical system². Distinguishing the more general class of coloured (autocorrelated) noise has proven more difficult because, unlike additive noise, the correlation between predicted and actual values measured may decrease with time—a property synonymous with chaos. Here, we show that by determining the scaling properties of the prediction error as a function of time, we can use nonlinear prediction to distinguish between chaos and random fractal sequences. Random fractal sequences are a particular class of coloured noise which represent stochastic (infinite-dimensional) systems with power-law spectra. Such sequences have been known to fool other procedures for identifying chaotic behaviour in natural time series⁹, particularly when the data sets are small. The recognition of this type of noise is of practical importance, as measurements from a variety of dynamical systems (such as three-dimensional turbulence, two-dimensional and geostrophic turbulence, internal ocean waves, sandpile models, drifter trajectories in large-scale flows, the motion of a classical electron in a crystal and other low-dimensional systems) may over some range of frequencies exhibit power-law spectra.

In the past, the identification of chaotic dynamics from time series relied heavily on the estimation of the dimension of the underlying attractor³⁻⁶, but this approach has certain drawbacks. First, the algorithms involved^{7,8} require a large number of data points, which are often not available. Second, even when enough points are available, a finite dimension may not be indicative of deterministic chaos. For small data sets, random fractal sequences (fractional brownian motions) may show an anomalous scaling which can be interpreted as a finite dimension^{9,10}. Fractional brownian motions (FBM) are random processes and thus are dictated by an infinite number of degrees of freedom. They have power spectra of the form $P(f) = Cf^{-a}$, where $a = 2H + 1$ with $0 < H < 1$. The exponent H is related to the fractal dimension, D_f , of the trail of the fractional brownian motion through $D_f = 1/H$. Irrespective of the relevance of such sequences to the natural world, these drawbacks mean that the methods have often required subjective judgement about whether an attractor of a given dimension exists, and other approaches have been sought.

One of these approaches is nonlinear prediction¹¹⁻¹³ which is becoming indispensable in the study of chaotic dynamical systems. The idea behind using nonlinear prediction as a signature of chaos is simple. Chaotic systems obey certain rules. The limited predictive power of chaotic dynamical systems is because they are sensitive to initial conditions and because we cannot have perfect measurements (which require an infinite

Received 6 April; accepted 18 June 1992.

- Paul, J. et al. in *Gamma Ray Line Astrophysics AIP Conf. Proc.* 232 (ed. Durouchoux, Ph. & Prantzos, N.) 17-28 (1991).
- Bouchet, L. et al. *Astrophys. J.* **383**, L45-L48 (1991).
- Lingenfelter, R. E. & Ramaty, R. in *The Center of the Galaxy* (ed. Morris, M.) 587-605 (Kluwer, Dordrecht, 1989).
- Sunyaev, R. A. et al. *Astr. Astrophys.* **247**, L29-L32 (1991).
- Bally, J. & Leventhal, M. *Nature* **353**, 234-237 (1991).
- Mirabel, I. F., Morris, M., Wink, J., Paul, J. & Cordier, B. *Astr. Astrophys.* **251**, L43-L46 (1991).
- Skinner, G. K. et al. *Astr. Astrophys.* **252**, 172-178 (1991).
- Condon, J. J. *Astrophys. J.* **287**, 461-474 (1984).
- Cordier, B. et al. *Astr. Astrophys. Suppl.* (in the press).
- Sunyaev, R. A. et al. *Sov. Astr.* **17**, 54-57 (1991).
- Woodworth, A. W., Higgs, L. A. & Gregory, P. C. *Astr. Astrophys.* **84**, 379-382 (1980).
- Fomalont, E. B. & Geldzahler, B. J. *Astrophys. J.* **383**, 289-294 (1991).

amount of information). One would therefore expect chaos to be characterized by a decrease of the correlation between predicted and actual values as prediction time increases. This property can be used to differentiate between chaos and additive uncorrelated noise. Additive noise produces a fixed amount of error regardless of the prediction time, as has been demonstrated¹ for deterministic maps and data from the natural world (the same conclusions were reached by Wolpert and Miall¹⁴ using the same data and a neural network as the prediction method). If a system is chaotic, then the decrease of predictive power with prediction time is equivalent to the presence of a positive Lyapunov exponent. The decrease with time of the correlation between predicted and actual values can indeed be used to infer the largest positive Lyapunov exponent². Those correlation functions do not require extremely large samples and have thus become a popular alternative to dimension estimates.

But how does nonlinear prediction perform when it comes to random fractal sequences? If random fractal sequences 'fool' the algorithms that characterize the structure of chaotic attractors, they may also fool prediction approaches whose signatures are based on the existence of those structures. In fact, as we will show, the correlation between predicted and actual values of fractional brownian motions also decreases with prediction time. In this case nonlinear prediction may have the same drawback as dimension estimates, and therefore may not be a definitive way of identifying chaos. Sugihara and May¹ speculated that the relationship between correlation and prediction time might scale differently but did not offer a solution to this central problem. Here we present theoretical arguments together with computer simulations that settle this issue.

If the predicted value at some time t is assumed to be a random variable x_t and the actual value a random variable y_t , then the Pearson's correlation coefficient between these two distributions, $r(t)$, ranges in magnitude between zero and one for uncorrelated and identical distributions, respectively. The correlation coefficient is defined as

$$r(t) = (\langle x_t y_t \rangle - \langle x_t \rangle \langle y_t \rangle) / \sigma(x_t) \sigma(y_t) \quad (1)$$

where the triangular brackets denote the average over a series of predictions and σ denotes the standard deviation. For stationary chaotic signals the correlation can take the form²

$$r(t) = 1 - (s^2(0) e^{2Kt}) / 2\sigma^2(y_t) \quad (2)$$

where $s(0)$, $\sigma(y_t)$ and K are positive constants. For stationary

processes $\sigma(y_t)$ is considered independent of the prediction time, t . Equation (2) dictates that the correlation should decrease exponentially with prediction time. Chaotic systems generally have positive Lyapunov exponents and exponential divergence of nearby trajectories.

If x is a FBM then $x_t = N(0, \propto t^{2H})$ —that is, it is normally distributed with zero mean and a variance proportional to t^{2H} —and $\text{cov}(x_t, y_t) = \langle x_t y_t \rangle$. In nonlinear prediction it is common to consider that

$$y_t = f\left(\sum_i w_i x_{t-i}\right) = f(z_t) \quad (3)$$

where f is some nonlinear function ranging between zero and one, and w_i are coefficients. If we expand f into a Taylor series we get

$$\begin{aligned} \langle x_t y_t \rangle &= \text{cov}(x_t, y_t) \\ &= \text{cov}(x_t, f(0) + z_t f'(0) + z_t^2 f''(0)/2 + \dots) \\ &= \text{cov}(x_t, f(0)) + f'(0) \text{cov}(x_t, z_t) \\ &\quad + f''(0) \text{cov}(x_t, z_t^2)/2 + \dots \end{aligned} \quad (4)$$

where $f(0)$, $f'(0)$ and $f''(0)$ represent some constants. If we consider only terms of second order or lower, then according to the problem in hand we have $\text{cov}(x_t, f(0)) = 0$ and $\text{cov}(x_t, z_t^2) = E(x_t z_t z_t) = 0$ (ref. 15). This allows us to write equation (4) as

$$\langle x_t y_t \rangle = A \text{cov}(x_t, z_t) \quad (5)$$

where $A = f'(0)$. Recalling that if $X = N(0, \min(t_1, t_2))$ then $\text{cov}(X_{t_1}, X_{t_2}) = \min(t_1, t_2)$ (ref. 15), and considering equation (3) we have

$$\begin{aligned} \langle x_t y_t \rangle &= A \text{cov}\left(x_t, \sum_i w_i x_{t-i}\right) \\ &= A \sum_i w_i \text{cov}(x_t, x_{t-i}) \\ &= A \sum_i w_i \min[t^{2H}, (t-i)^{2H}] \\ &= A \sum_i w_i (t-i)^{2H} \end{aligned} \quad (6)$$

Consequently, the correlation coefficient takes the form

$$r(t) = B \sum_i w_i (1 - i/t)^{2H} \quad (7)$$

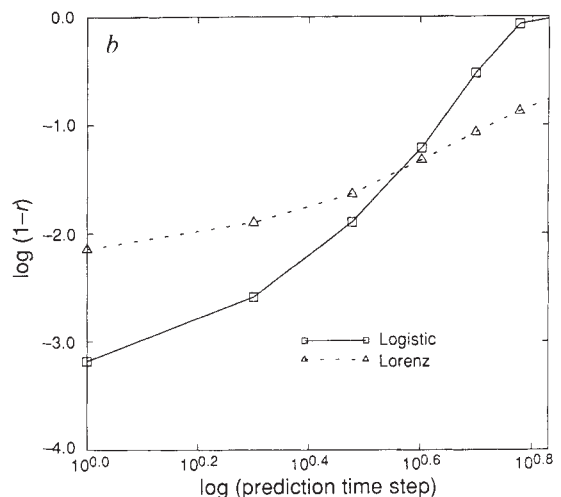
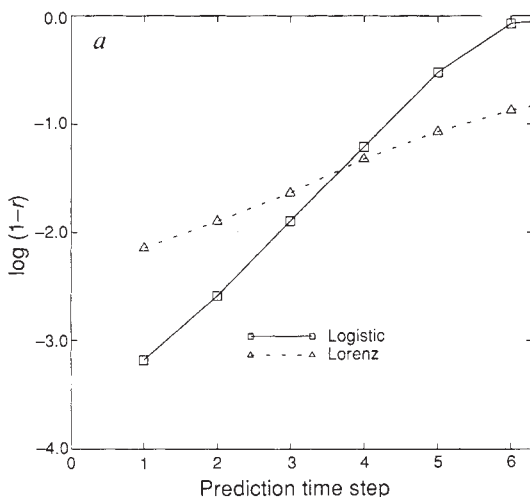


FIG. 1 Nonlinear prediction can be used to distinguish chaotic signals from random fractal sequences. *a*, Logarithm of $1 - r(t)$ against prediction time step, t ; *b*, logarithm of $1 - r(t)$ against $\log t$ for single realizations of the x coordinate of the Lorenz system, and the logistic map. These simulations

show the expected scaling from the theoretical arguments developed in the text. For short prediction times, scaling is observed in the semi-log plot, indicating that the curve of $1 - r(t)$ against t curve for the chaotic signals is exponential.

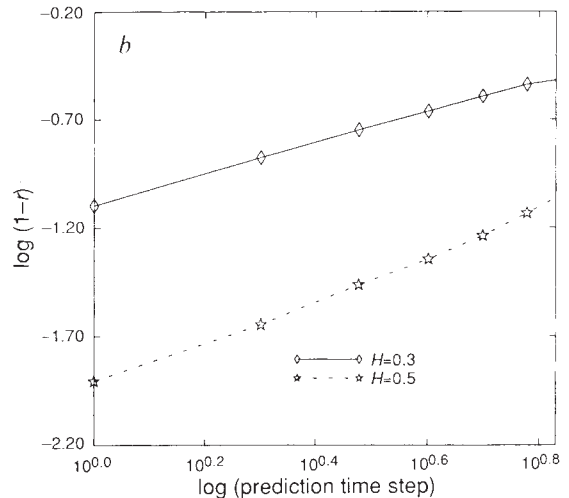
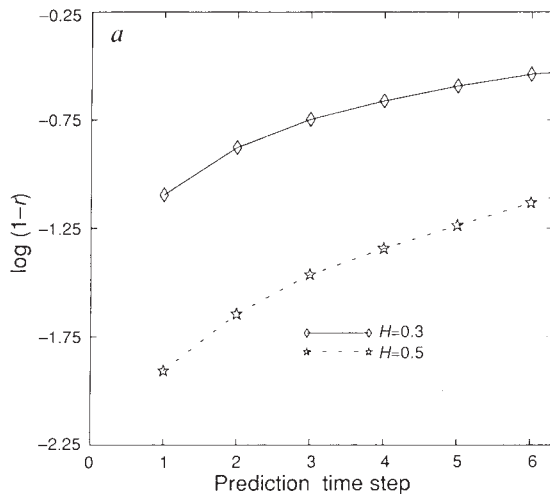


FIG. 2 As Fig. 1, but showing single realization results for two FBMs with $H=0.3$ and 0.5 . Here scaling is observed in the log-log plot indicating that

for the FBMs the curve of $1-r(t)$ against t is a power-law curve, as expected.

where B is some constant. Considering that in our case $\text{cov}(x_t, z_t^2) = E(x_t z_t z_t) = 3 \text{cov}(x_t, z_t) \text{Var}(z_t)$ and that $\text{cov}(x_t, z_t^4) = E(x_t z_t z_t z_t) = 0$, addition of the next two higher-order terms in the expansion modifies equation (7) to

$$r(t) = \left[B + C \left(\sum_i w_i^2 (t-i)^{2H} + \sum_{i < j} w_i w_j (t-j)^{2H} \right) \right] \times \sum_i w_i (1-i/t)^{2H} \quad (8)$$

where C is some other constant. Equations (7) and (8) do not include any exponential terms. They represent curves determined by a synthesis of power-law terms. This is a direct consequence of the power-law characterization of FBM. Including additional terms in the expansion of f does not change this property. The above arguments, and especially equations (2) and (7) or (8), show that we should expect differences in the scaling of the correlation coefficient with prediction time between chaotic signals and FBMs. For chaotic systems the logarithm of $1-r(t)$ should be a linear function of prediction time step t , and for FBMs it should be a linear function of $\log t$.

These scaling differences are emphasized by computer simulations. Figure 1a is a plot of $\log(1-r(t))$ against t , and Fig. 1b

is a plot of $\log(1-r(t))$ against $\log t$ showing single realization results for the x coordinate of the Lorenz system, and for the logistic map $x_{n+1} = 4x_n(1-x_n)$. The integration step for the Lorenz system was 0.03. The prediction method was similar to that used in ref. 1, which is a simple variant of the Farmer and Sidorowich interpolative approach¹¹. For each example we take 1,000 values from the corresponding time series, use the first 500 as a database, and make predictions on the last 500 values. Correlation coefficients are thus based on sample sizes of ~ 500 . Note that for the Lorenz time series, the prediction time step is the integration step. Figure 2 is similar to Fig. 1 but shows single realization results for two FBMs with $H=0.3$ and 0.5 . In our simulations the embedding dimension is always $2D_f+1$. The time delay used to define the coordinates of the embedding space was $\tau=1$. The results show a decrease of the correlation coefficient with prediction time in all cases. In accordance with our theoretical arguments, however, Figs 1 and 2 indicate that for short prediction times, $\log(1-r(t))$ is a linear function of t for the chaotic signals and a linear function of $\log t$ for the FBMs. Although the role of τ in nonlinear prediction needs to be better understood, preliminary experiments with different values of τ indicated that the expected scaling is preserved; other prediction techniques, such as neural networks¹³, give

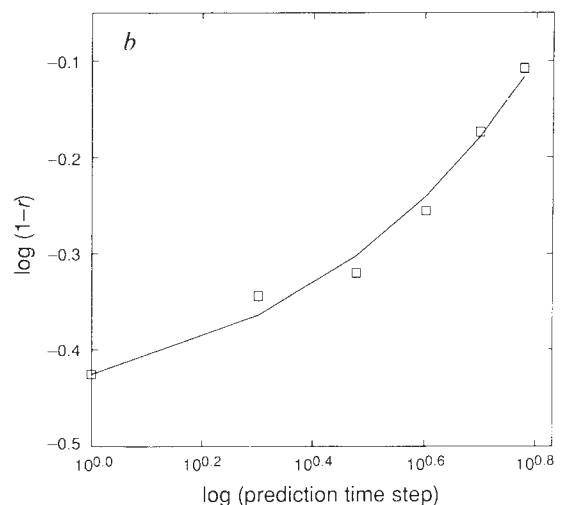
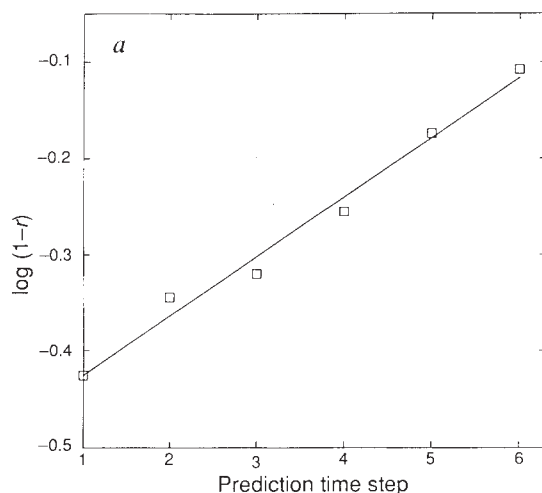


FIG. 3 Nonlinear prediction on a time series of the Southern Oscillation Index. a, $\log(1-r(t))$ against t ; b, $\log(1-r(t))$ against $\log t$. For short prediction time steps, scaling is observed in the semi-log plot indicating

that the curve of $1-r(t)$ against t is exponential, and thus the time series is a chaotic signal.

similar results but seemed to yield less accurate predictions than the interpolative approach.

For periodic signals with additive uncorrelated noise, the correlation coefficient is independent of the prediction time. A fall in the correlation with prediction time does not indicate a periodic signal with additive noise, but it may indicate a chaotic signal with or without noise (in the chaotic case, additive noise produces a constant offset in the correlation which would not obliterate the fall-off). Our work provides the means to go one step further and decide whether such a fall-off is indicative of a random fractal sequence. A direct approach will be to check the scaling behaviour of $\log(1-r(t))$ against t and of $\log(1-r(t))$ against $\log t$. For example, Fig. 3 shows results from a time series of the Southern Oscillation Index, which is derived from the mean sea-level pressure difference between the Tahiti and Darwin stations. The record is 1,248 values long (monthly values from January 1883 to December 1986) and is related to the El Niño which is hypothesized to be chaotic¹⁶. A dimension of around five has been suggested for this type of data¹⁷. Figure 3a is a semi-log plot and Fig. 3b a log-log plot of $1-r(t)$ against t . The solid lines indicate best fits. For the semi-log plot the best fit is linear, whereas for the log-log plot it appears to be non-linear. According to our theory this indicates that the time series is indeed chaotic. In some cases, noise, data imperfections and data length may make the identification of the actual scaling

difficult. In such cases nonlinear prediction could be used statistically. For example, if a time series is suspected to be a FBM, then prediction results could be compared with the average prediction properties of the family of FBMs in question. □

Received 12 November 1991; accepted 5 May 1992.

1. Sugihara, G. & May, R. M. *Nature* **344**, 734-741 (1990).
2. Wales, D. J. *Nature* **350**, 485-488 (1991).
3. Tsonis, A. A. & Elsner, J. B. *Nature* **333**, 545-547 (1988).
4. Essex, C., Lookman, T. & Nerenberg, M. A. H. *Nature* **326**, 64-66 (1987).
5. Frank, G. W., Lookman, T., Essex, C. & Nerenberg, M. A. H. *Physica D* **46**, 427-438 (1990).
6. Schaffer, W. M., Olsen, L. F., Truty, G. L., Fulmer, S. L. & Graser, D. J. in *From Chemical to Biological Organization* (eds Marcus, M., Muller, S. C. & Nicolis, G.) (Springer, New York, 1988).
7. Grassberger, P. & Procaccia, I. *Physica D* **9**, 186-208 (1983).
8. Grassberger, P. & Procaccia, I. *Phys. Rev. Lett.* **50**, 346-349 (1983).
9. Osborne, A. R. & Provenzale, A. *Physica D* **35**, 357-381 (1989).
10. Theiler, J. *Phys. Lett. A* **155**, 480-492 (1991).
11. Farmer, J. & Sidorowich, J. J. *Phys. Rev. Lett.* **62**, 845-848 (1987).
12. Casdagli, M. *Physica D* **35**, 335-356 (1989).
13. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. *Nature* **323**, 533-536 (1986).
14. Wolpert, D. M. & Miall, R. C. *Proc. R. Soc. B* **242**, 82-86 (1990).
15. Anderson, T. W. *An Introduction to Multivariate Statistics* (Wiley, New York, 1958).
16. Valis, G. K. *Science* **232**, 243-245 (1986).
17. Hense, A. *Beitr. Phys. Atm.* **60**, 34-47 (1987).

ACKNOWLEDGEMENTS. We thank J. D. Farmer, K. P. Georgakakos and J. H. Beder for discussion and comments, S. Nicholson for providing the Southern Oscillation Index data, and J. J. Sidorowich for his review and advice. Parts of this work were supported by the NOAA TOGA programme on prediction.

Large-scale synthesis of carbon nanotubes

T. W. Ebbesen & P. M. Ajayan

Fundamental Research Laboratories, NEC Corporation, 34 Miyukigaoka, Tsukuba 305, Japan

INTEREST in carbon fibres^{1,2} has been stimulated greatly by the recent discovery of hollow graphitic tubules of nanometre dimensions³. There has been much speculation about the properties and potential application of these nanotubes⁴⁻⁸. Theoretical studies predict that their electronic properties will depend on their diameter and degree of helicity^{4,5}. Experimental tests of these ideas has been hampered, however, by the lack of macroscopic quantities of the material. Here we report the synthesis of graphitic nanotubes in gram quantities. We use a variant of the standard arc-discharge technique for fullerene synthesis under a helium atmosphere. Under certain conditions, a carbonaceous deposit forms on one of the graphite rods, consisting of a macroscopic (diameter of about 5 mm) cylinder in which the core comprises pure nanotubes and nanoscale particles in high yield. The purity and yield depend sensitively on the gas pressure in the reaction vessel. Preliminary measurements of the conductivity of the bulk nanotube material indicate a conductivity of about 100 S cm^{-1} .

The carbon arc experiments are done in a reaction vessel through which an inert gas (He, Ar and so on) flows at a controlled pressure. We apply a potential of $\sim 18 \text{ V}$, either a.c. or d.c., between two graphite rods (one 6 mm and the other 9 mm in diameter) in the vessel. As the rods are brought close together a discharge occurs resulting in the formation of a plasma. As the smaller rod is consumed, the rods are kept at a constant distance from each other of $\sim 1 \text{ mm}$. On the larger rod, a carbonaceous deposit forms which may contain nanotubes under the right conditions. The electric current depends on the size of the rods, their separation, the gas pressure and so on, but typically it is $\sim 100 \text{ A}$. It was while trying to prepare derivatives of fullerenes under helium atmosphere and alternating current that we discovered how to make nanotubes in reasonable quantities. The techniques were then optimized as described below.

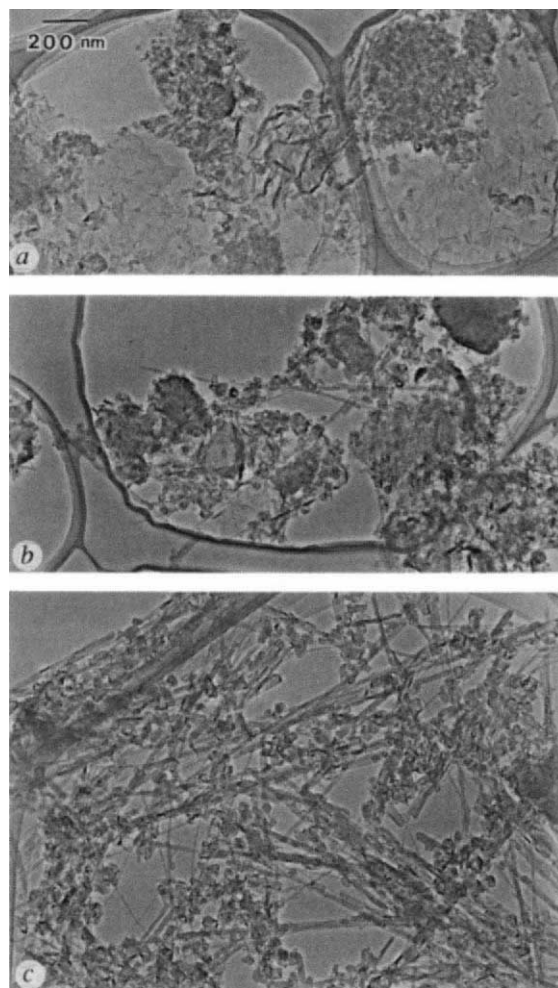


FIG. 1 Low-magnification electron micrographs of samples prepared in a helium atmosphere at pressures (a) 20 torr, (b) 100 torr and (c) 500 torr. There is a striking increase in the number of tubes as the pressure is increased.

To optimize the yield of nanotubes, we varied conditions such as the type of inert gas, the nature of the current (a.c. or d.c.), the voltage and the relative rod size. The relative yield of nanotubes was monitored by transmission electron microscopy after grinding the carbonaceous deposit and sonicating it in ethanol. Figure 1 shows the effect of helium pressure on the yield of nanotubes at 18 V d.c. At 20 torr, hardly any nanotubes can be found. Only amorphous, sheet-like and glassy graphitic material is formed. At 100 torr (Fig. 1b), a few nanotubes appear in the sample. At 500 torr (Fig. 1c) the entire sample consists of carbon nanotubes and nanoparticles (cage structures). We tested the effect of increasing the pressure to 2,500 torr, but the quality of the sample did not change significantly. The amount of recovered material, however, did decrease. Therefore at ~ 500 torr the total yield of nanotubes as a proportion of graphitic starting material is optimal. Under those conditions, $\sim 75\%$ of the consumed graphite rod is converted to a deposit of cylindrical shape and similar diameter.

Figure 2 shows the typical rod-shaped deposit together with the large 9-mm graphite rod on which it forms. The rate at which the rod deposit forms depends on the gas pressure; in the best conditions for producing nanotubes, it is a few millimetres per minute. Two distinct regions can be seen in the cross-section (Fig. 2, inset): an inner black core made only of nanotubes and nanoparticles, and an outer grey metallic hard shell in which no nanotubes are seen by transmission electron microscopy. In other words, the deposit forms a macroscopic tubular structure packed with nanotubes. This preparation method is therefore convenient, as the core can easily be separated from the shell. The weight ratio of the core to the shell is ~ 1 to 2, corresponding to a nanotube-to-particle yield of 25% of the starting material. This compares favourably to the yield for fullerenes, typically a few per cent of the starting material and 8% of the recovered soot. Within the sample the ratio of tubes to particles is at least 2 to 1 by weight; treating the sample in an ultrasonic bath can free many tubes from the particles which are originally stuck together. The deposit containing the nanotubes is pitch black, but grinding and smearing it over a surface produces a shiny greyish metallic lustre.

We can also prepare nanotubes using alternating current, or in argon (d.c. or a.c.), but the total yields are much smaller. The reason for this might be that the nanotubes are only formed during one part of the a.c. cycle, so the yield is reduced. The inert gas pressure remains critical in all cases and should be ~ 500 torr or larger. We varied the applied potential between 10 V and 18 V, but found that it has little effect on the quality of the sample provided that a plasma is formed.

The carbon nanotubes have the structures reported by Iijima³ and typically consist of two or more concentric shells of carbon sheets (Fig. 3). The diameter of the nanotubes in our samples ranges between 2 and 20 nm, whereas the lengths are several micrometres. The tips of the tubes are capped with pentagons, giving unique angular relationships between the tube surfaces

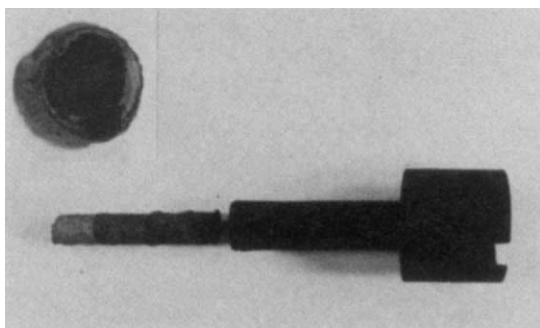


FIG. 2 Rod-shaped deposit, together with the larger graphite rod on which it forms. Inset: cross-section of the deposit (diameter ~ 5 mm).

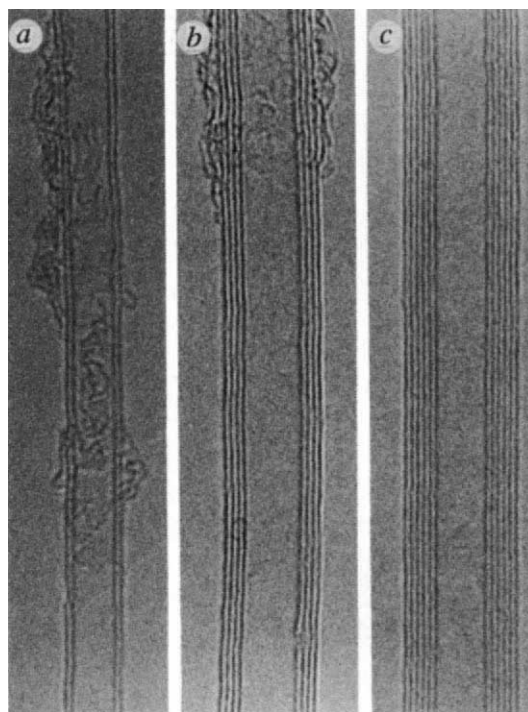


FIG. 3 High-resolution electron micrographs of typical tubes present in samples prepared under helium at 500 torr as described in the text. The number of layers ranges from two to many. The distance between consecutive carbon layers in the tubes is 0.34 nm.

and the tips⁹, and a stacked morphology is often seen for the inner structure.

The core material containing the nanotubes is macroscopically fibrous, with the fibres parallel to the direction of the current, suggesting that the nanotubes themselves might be aligned in the sample. Preliminary scanning electron microscope studies support this observation. The fact that nanotubes are found only in the core of the sample and not on the colder edge indicates that the temperature and cooling rate are critical for their formation. This is supported by the importance of the inert gas pressure in the chamber. We also observed that if the plasma is stable, a more even sample is produced. This is similar to the situation of vapour-grown crystals and suggests that with accurate control of the plasma temperature and its uniformity, the diameter and length of the nanotubes could be more precisely controlled. Another possible way of narrowing the distribution of nanotubes is by separating them after formation. This seems to be more difficult than purifying C_{60} from larger fullerenes, when standard chromatography methods can be used. Separation of the nanotubes from the nanoparticles is easier because of the difference in size and shape.

Theoretical calculations predict that a tubular structure is energetically more favourable than a flat sheet of carbon hexagons of width equal to the circumference⁶. The fact that we observe no small sheets or sheet aggregates in our pure samples supports this. Other calculations on the electronic structure of nanotubes predict semiconductor or metallic properties depending on their diameter and helicity^{4,5}. Although we have not yet been able to measure the properties of single tubes, the bulk material is indeed conducting with a conductivity larger than 100 S cm^{-1} . Considering the discontinuity and diversity in the bulk structure, some tubes must have a much higher conductivity than this average value.

The ability to mass-produce nanometre fibrous materials with large aspect ratios (length/diameter), such as the carbon nanotubes, is of importance in materials science¹⁰. Our

large-scale synthesis method for nanotubes is simple and easily accessible to others with an interest in this novel material, and should enable studies to be undertaken to evaluate their physical properties and assess potential applications in areas such as catalysis, composite materials and nanowires. □

Received 15 June; accepted 26 June 1992.

1. Speck, J.S., Endo, M. & Dresselhaus, M. S. *J. Cryst. Growth* **94**, 834–848 (1989).
2. Tibbitts, G. G. *J. Cryst. Growth* **66**, 632–638 (1984).
3. Iijima, S. *Nature* **354**, 56–58 (1991).
4. Mintmire, J. W., Dunlap, B. I. & White, C. T. *Phys. Rev. Lett.* **68**, 631 (1992).
5. Hamada, N., Sawada, S. & Oshiyama, A. *Phys. Rev. Lett.* **68**, 1579–1581 (1992).
6. Sawada, S. & Hamada, N. *Solid State Commun.* (in the press).
7. Robertson, D. H., Brenner, D. W. & Mintmire, J. W. *Phys. Rev.* (in the press).
8. Adams, G. B. *et al. Science* (in the press).
9. Iijima, S., Ichihashi, T. & Ando, Y. *Nature* **356**, 776–778 (1992).
10. Calvert, P. *Nature* **357**, 365–366 (1992).

ACKNOWLEDGEMENTS. We thank H. Hiura for assistance.

Negative Poisson ratios in crystalline SiO₂ from first-principles calculations

Nitin R. Keskar & James R. Chelikowsky

Department of Chemical Engineering and Materials Science, Minnesota Supercomputer Institute, University of Minnesota, Minneapolis, Minnesota 55455, USA

THE Poisson ratio of a solid characterizes its response to uniaxial stress. It is defined as the negative ratio of the transverse strain to the corresponding axial strain. Normally, this ratio is positive, as most solids expand in the transverse direction when subjected to a uniaxial compression. Although a negative Poisson ratio is not forbidden by thermodynamics, it is rare in crystalline solids: the results of recent experiments¹ which observed a negative Poisson ratio in α -cristobalite were therefore unexpected. We have investigated the elastic behaviour of α -cristobalite and other forms of silica with first-principles calculations and classical interatomic potentials. Our calculations reproduce the negative Poisson ratio in α -cristobalite, and predict that α -quartz, the most common form of crystalline silica, will also exhibit a negative Poisson ratio under large uniaxial tension. We attribute the occurrence of a negative Poisson ratio in low-density silica polymorphs to the high rigidity of the SiO₄ tetrahedra.

For a crystal with a tetragonal or hexagonal unit cell defined by the lattice parameters a and c , the Poisson ratio for a uniaxial loading along the c -axis is given by

$$\sigma = -\frac{\Delta a/a}{\Delta c/c} = -\frac{\Delta \ln(a)}{\Delta \ln(c)} \quad (1)$$

Solids under such conditions usually deform to conserve volume: compression in one direction results in an expansion in the transverse direction. This gives a positive Poisson ratio. An empirical rule of $\sigma > 0$ is satisfied by virtually all known solids. A negative value of Poisson ratio is not forbidden: the requirement that the strain energy for an elastic isotropic solid be non-negative leads only to the restriction $-1 \leq \sigma \leq \frac{1}{2}$. For an anisotropic solid, the allowed range is even wider². Two examples of metallic crystal systems with negative Poisson ratios are known^{3,4}. It is difficult to measure a negative Poisson ratio, and the observation of such a ratio can be spurious⁵. A theoretical confirmation can validate experimental results, but such calculations have been rare.

Weidner *et al.*¹ have recently measured single-crystal elastic moduli of α -cristobalite. The polycrystalline aggregate values of the moduli were: bulk modulus $B = 16.4$ GPa; shear modulus $G = 39.06$ GPa. These values result in a Poisson ratio of -0.16

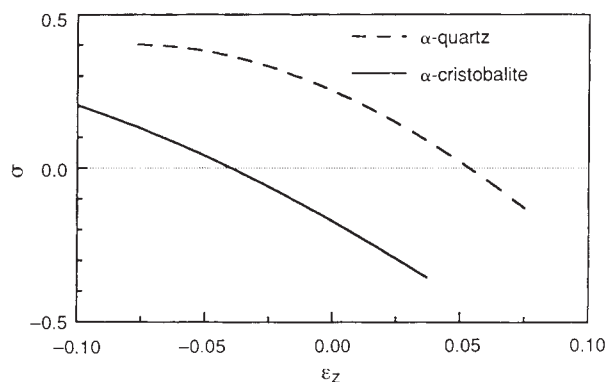


FIG. 1 Variation of the Poisson ratios of α -quartz and α -cristobalite as a function of uniaxial strain along the c -axis. This figure is based on our pair-potential calculations. Note that a positive strain occurs for a material under tension.

for α -cristobalite according to the following relationship for isotropic materials:

$$\sigma = \frac{3B - 2G}{2(3B + G)} \quad (2)$$

This result was certainly counterintuitive. α -cristobalite is a covalent polymorph of silica in which the silicon cation is fourfold coordinated. It has the same basic building blocks, SiO₄ tetrahedra, as other common forms of crystalline silica such as α - and β -quartz. The latter crystals are reported to have a positive Poisson ratio at ambient pressure⁶.

We analysed the Poisson ratio of crystalline silica using two theoretical approaches. One approach is based on classical interatomic potentials. The other is based on *ab initio* pseudopotentials and is fully quantum-mechanical. The classical pair potential that we have used in this study was developed using Hartree-Fock self-consistent calculations on SiO₄⁴⁻ clusters⁷. This interatomic potential seems to work well for structural and elastic properties of silica structures^{7,8}, but one might not expect it to work for details of the elastic properties of silica, especially for open structures where angular forces might be important. To verify the results we used quantum mechanical calculations. These calculations give accurate structural properties, but they are computationally intensive. We have used 'soft' pseudopotentials constructed within the local density approximation⁹. These pseudopotentials are widely applicable and transferable^{9,10}. Details of the methods can be found elsewhere^{11–13}.

Determining the Poisson ratio of α -cristobalite is complicated by its large unit cell and the number of degrees of freedom of the atoms. α -cristobalite has a tetragonal unit cell which is completely defined by lattice parameters a and c , and the internal coordinates u , x , y and z (ref. 14).

In calculating the Poisson ratio, we used the definition in equation (1). For a fixed value of c , we let the other lattice parameter and internal coordinates relax to find a minimum-energy structure. We then changed the values of c and let the internal coordinates and a relax to find the minimum-energy structure for the new value of c . Owing to the computational difficulties of this energy minimization, we optimized geometries using pseudopotential calculations at only four values of c . In our pair-potential calculations, we examined the structure at 50 values of the lattice parameter c . This procedure gives us the energy of the structure and the lattice constant a as a function of the lattice constant c for a material under a uniaxial stress.

In Fig. 1, we show the variation of the Poisson ratio as a function of the engineering strain along the c -axis, defined as $\epsilon_z = (c - c_0)/c_0$. For $\epsilon_z \geq -0.04$ the Poisson ratio assumes

negative values. But if the material is further compressed along the c -axis, the Poisson ratio becomes more and more positive. We predict that at high compressive strains the ratio will saturate at ~ 0.3 , as compared with the value of 0.5 for an incompressible solid. Our pair-potential and pseudopotential calculations give values -0.17 and -0.2 respectively for the Poisson ratio at the unstrained geometry, in good agreement with the experimental value of -0.16 . Our fully quantum mechanical calculations thus confirm the pair-potential calculations.

Given the ease of pair-potential calculations, we concentrated on this approach for α - and β -quartz. Comparison with these forms of silica, which have positive Poisson ratios at ambient conditions, should bring out the differences between these similarly coordinated structures. We optimized the structures of α - and β -quartz for a uniaxial loading along the c -axis. Over a wide range, $-0.15 < \epsilon_z < 0.15$, β -quartz has a Poisson ratio of 0.4, compared with the experimental value⁶ of 0.2.

At ambient pressure, we have computed a Poisson ratio of ~ 0.2 for α -quartz as compared with the experimental value⁶ of 0.08. For this structure, $\ln a$ does not behave linearly as a function of $\ln c$, so the Poisson ratio varies. Its variation as a function of strain shows similar trends for both α -quartz and α -cristobalite (Fig. 1). At a large compressive strain, both have a positive Poisson ratio, as expected for a solid that becomes increasingly incompressible. As the compressive strain is reduced, the Poisson ratio decreases, and for large tensile strain, both crystals have a negative ratio. The main difference between α -quartz and α -cristobalite is that the ratio becomes negative for α -cristobalite at ambient pressure whereas for α -quartz the ratio remains positive. On the basis of this work, we predict that the Poisson ratio for α -quartz will be negative at high tensile strains as it is for α -cristobalite at ambient pressure.

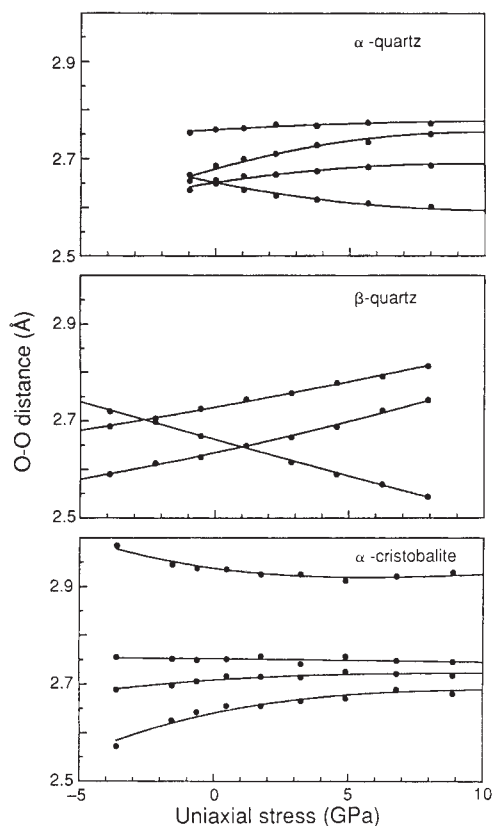


FIG. 2 Variation of the nearest-neighbour O-O distances in α - and β -quartz and α -cristobalite as a function of the uniaxial stress. We show only a few illustrative data points from our pair-potential calculations. The uniaxial stress σ_z is calculated as $\sigma_z = E_z \epsilon_z$. Young's modulus E_z is given by $E_z = 3B_0(1 - 2\sigma)$. B_0 is the bulk modulus from our calculations.

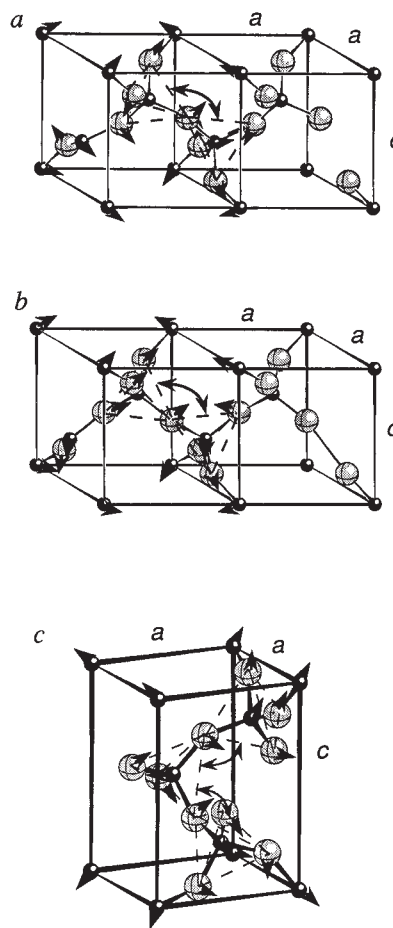


FIG. 3 Motions of atoms in (a) α -quartz, (b) β -quartz and (c) α -cristobalite as they are subjected to a uniaxial tension. Note that the displayed geometries are those at ambient pressure and that the motions of the atoms are relative to the centre of the corresponding unit cells. We also show the variation of the intertetrahedral angles. The dark and light circles are silicon and oxygen atoms, respectively. Lattice parameters a and c are labelled.

The microscopic origin of the negative Poisson ratio is subtle. An unanticipated conclusion is that angular forces are not essential to produce the negative Poisson ratio behaviour in silica. Pairwise forces alone, as determined from classical interatomic potentials, can reproduce the Poisson ratio in these materials. We therefore examined the nearest-neighbour interatomic distances Si-O and O-O. (We expect on physical grounds that the Si-Si distance should not be relevant). We found no significant differences or trends for the Si-O bonds for these polymorphs. The Si-O bond lengths vary by no more than 0.02 Å over a wide range of stresses. The O-O distances, however, show an interesting trend (Fig. 2). For uniaxial tension, the four unique O-O distances in α -quartz and α -cristobalite both tend to saturate to constant values. Constant O-O distances are synonymous with rigid SiO_4 tetrahedra as fixing the O-O distances, fixes the O-Si-O angles. This relationship between O-O distances and the tetrahedral angles may account for the surprising accuracy of pair potentials for silica.

To verify that the rigidity of the tetrahedral SiO_4 units causes the negative Poisson ratio in α -quartz and α -cristobalite, we adopted the following approach. In both these materials, there are four distinct O-O distances and a unique Si-O distance. We found expressions for these five distances in terms of the four internal coordinates and the lattice parameter a and fixed the distances at values occurring in crystals of these polymorphs. At different values of the c parameter, we solved for u , x , y , z

and a with the above constraint. This calculation is equivalent to compressing the materials along the c -axis while maintaining the rigidity of the tetrahedra. We plotted $\ln a$ as a function of $\ln c$ and calculated the Poisson ratios. We found that in these artificially constrained structures of rigid tetrahedra, the Poisson ratios of both α -quartz and α -cristobalite are -0.6 and -0.5 respectively. This confirms the role of rigid tetrahedra in imparting negative Poisson ratios to these polymorphs.

The motion of the tetrahedra is illustrated more explicitly in Fig. 3, where we have indicated the motions of atoms, as determined from our pair-potential calculations, in the unit cells of α - and β -quartz and α -cristobalite as they are subjected to a uniaxial tension. As the c -axis is extended, the SiO_4 units in α -cristobalite rotate 'outward' by increasing the Si-O-Si bridging angle. From our quantum mechanical calculations and from other computations¹⁵, this angle is known to be soft. The tetrahedral rotation of the SiO_4 units allows a to increase along with the volume of the unit cell, and a negative Poisson ratio occurs. This behaviour does not follow for high compressive strains because the SiO_4 unit eventually distorts, as indicated by the variation in the O-O distances in Fig. 2. α -cristobalite, and α -quartz under tension, are similar in that they correspond to low-density forms of silica. If sufficient distance occurs between

the tetrahedral units, then the strain can be accommodated by a reduction in the intertetrahedral distances while preserving the SiO_4 units. This situation seems to be necessary for a negative Poisson ratio. $\square$

Received 28 February; accepted 18 May 1992.

- Weidner, D. J., Yeganeh-Haeri, A. & Parise, J. *Eos Trans.* **71**, 619 (1990).
- Lempriere, B. M. *Am. Inst. Aeronaut. Astronaut. J.* **6**, 2226-2227 (1968).
- Love, A. E. H. *A Treatise on the Mathematical Theory of Elasticity*, 4th Edn 163 (Dover, New York, 1944).
- Li, Y. *Phys. Stat. Sol.* **38**, 171-175 (1976).
- Grimvall, G. *Thermophysical Properties of Materials*, 62 (North-Holland, Amsterdam, 1986).
- Landolt-Börnstein: *Numerical Data and Functional Relationships in Science and Technology*, new series V/1b (ed. Angenheister, G.) 33 (Springer, Berlin, 1982).
- Tsuneyuki, S., Tsukada, M., Aoki, H. & Matsui, Y. *Phys. Rev. Lett.* **61**, 869-872 (1988).
- Keskar, N. R. & Chelikowsky, J. R. *Phys. Rev. B* (in the press).
- Troullier, N. & Martins, J. L. *Solid State Commun.* **74**, 613-616 (1990).
- Troullier, N. & Martins, J. L. *Phys. Rev. B* **43**, 1993-2006 (1991).
- Keskar, N. R., Troullier, N., Martins, J. L. & Chelikowsky, J. R. *Phys. Rev. B* **44**, 4081-4088 (1991).
- Chelikowsky, J. R., King, H. E. Jr, Troullier, N., Martins, J. L. & Glinnemann, J. *Phys. Rev. Lett.* **65**, 3309-3312 (1990).
- Chelikowsky, J. R., Troullier, N., Martins, J. L. & King, H. E. Jr. *Phys. Rev. B* **44**, 489-497 (1991).
- Wyckoff, R. W. G. *Crystal Structures*, 2nd edn, Vol. 1, 316-318 (Interscience, New York, 1964).
- Stixrude, L. & Bukowski, M. S. T. *Phys. Chem. Miner.* **16**, 199-206 (1988).

ACKNOWLEDGEMENTS. We thank H. E. King Jr for discussions. This work was supported by the Division of Materials Research, Office of Basic Energy Sciences, U.S. Department of Energy and by the Minnesota Supercomputer Institute.

Variability in sea-ice thickness over the North Pole from 1977 to 1990

A. S. McLaren*, J. E. Walsh†, R. H. Bourke‡, R. L. Weaver§ & W. Wittmann||

* Lamont-Doherty Observatory, Columbia University, Palisades, New York 10964 and Science Service, Inc., Washington DC, 20036, USA

† Department of Atmospheric Sciences, University of Illinois, Champaign-Urbana, Illinois 61801, USA

‡ Department of Oceanography, Naval Postgraduate School, Monterey, California 93943, USA

§ Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado 80309, USA

CHANGES in the thickness of polar sea-ice have the potential to provide a signal of climate change, but attempts to identify trends must take into account the range of natural variability. Here we present an analysis of measurements of the subsurface ice thickness (draft) of sea-ice around the North Pole made from 1977 to 1990. These data were collected during six submarine cruises in late April/early May of 1977, 1979, 1986, 1987, 1988 and 1990, and represent the most extensive dataset so far for ice draft in the central Arctic at the same season and location. The results reveal considerable interannual variability both in mean ice draft (± 1.0 m) and in open-water extent ($\pm 2.5\%$). This variability limits the confidence that can be placed in any apparent trends observed for sea-ice thickness or type since the late 1970s, and illustrates the need for a reliable baseline against which to assess future trends.

The Arctic Ocean is generally regarded as a potentially sensitive indicator of global change, especially as the predicted result of continued greenhouse-gas emissions is a global warming with high-latitude amplification¹. Such a warming could thus have important consequences for the thickness and extent of Arctic sea-ice. The resultant melting of the ice would presumably lead to a decrease in both its thickness and areal extent, as well as to an increase in the number of open or thin-ice areas within the Arctic ice pack. Arctic ice thickness indeed decreases con-

siderably as atmospheric CO_2 increases in simulations with coupled ocean-atmosphere models².

At present, most measurements of sea-ice thickness are based on submarine under-ice sonar data which record the ice draft, that is the submerged portion below sea level. True ice thickness includes the above-water freeboard which can be estimated theoretically or empirically. Bourke and Garrett³ used unclassified submarine cruise logs and sonar data to develop seasonal climatologies of ice draft in the Arctic Basin. Their compilation contains no reference to variability. The ice draft recorded by the submarines *Nautilus* (1958) and *Queenfish* (1970) along the same transect through the Canada Basin and over the North Pole during early August was 0.7-0.8 m thinner in 1970 than in 1958, and the areal coverage of open water and thin ice was greater in 1970 than in 1958 (ref. 4). Submarine sonar measurements indicated significantly thinner sea ice north of Greenland in 1987 than in 1976 (ref. 5).

The findings of McLaren⁴ and Wadhams⁵ raise intriguing questions about recent variations of sea ice in the context of climate change. From the evidence so far available, McLaren *et al.*⁶ concluded that too little is known about the natural variability of Arctic sea-ice thickness to draw any conclusions about climate-induced changes in the thickness based on comparative analyses of only a few sea-ice draft datasets collected within the Arctic Basin. Recently, however, Gloersen and Campbell⁷ reported that satellite-borne scanning multispectral microwave radiometer (SMMR) data collected by Nimbus 7 between October 1978 and August 1987 reveals significant decreases of ice extent and open water areas within the Arctic. Unfortunately, Nimbus 7 provides no SMMR data on ice extent and open water north of 84°N (ref. 8), thereby precluding findings concerning the portion of the Arctic Basin around the North Pole.

In an effort to learn more about the natural variability of sea-ice draft and the extent of open water or thin ice in the interior of the Arctic Basin, we recently completed an analysis of sea-ice data collected during six U.S. nuclear submarine transpolar cruises that took place in late April/early May of 1977, 1979, 1986, 1987, 1988 and 1990. All six submarines acoustically recorded the under-ice distribution of sea-ice draft at the North Pole and its immediate vicinity.

Over 200,000 km of sea-ice draft data have been collected within the Arctic basin by submarines since 1958. Less than 15% of this data has so far been analysed. The six submarine cruises

|| Deceased.

reported herein were chosen on the basis of data quality and because they are essentially coincident in time and location. Further, they span a 13-year period that essentially coincides with the availability of satellite data on the areal extent of Arctic Basin sea ice. The region surrounding the North Pole was chosen for this analysis because it has been repeatedly surveyed by U.S. nuclear submarines since 1958. No other portion of the Arctic basin has been subjected to such continuous sea-ice measurement, so we have focused on this particular area for our analysis of observed variability. This is not considered a severe limitation, however, as regions near the ice margins, which undergo significant secular changes, may indicate trends that are very local or short lived. Additionally, the sea-ice draft distribution within the central polar pack represents an integrated effect due to the advection of ice from many locales. Discernable trends measured here are therefore more likely to be representative of a much broader area.

Submarine sonar measurements of ice draft are continuously recorded for operational purposes. The raw data can thus contain a variety of sensor and transmission system errors. The data used here was provided by the U.S. Navy's Arctic Submarine Laboratory, San Diego, after careful editing by their staff to remove these errors through frequent calibration to open water/new ice indications from coincidentally recorded analog records. The resultant accuracy of the edited data is estimated to be ± 0.3 m (T. Luallin, personal communication). When averaged over 50–100-km segments, the errors associated with mean statistics are estimated to be ± 0.15 m.

The results of an in-depth statistical analysis of these data, using methodology described in refs. 4, 9, are summarized in Table 1 for 50-km and 100-km segments centred directly over the North Pole. Figures 1 and 2 show the mean drafts and open-water fractions for the 100-km segment. For our present purposes, ice with draft less than 3.0 m is assumed to be 'first-year

TABLE 1 Comparative sea-ice statistics for 50- and 100-km-long transects centred over the North Pole

	Flying fish 1977	Archer fish 1979	Archer fish 1986	Bill fish 1987	Archer fish 1988	Gurnard 1990
Ice draft						
Mean draft (m)						
50 km	4.2	4.0	2.8	4.1	3.3	3.6
100 km	4.0	4.0	3.1	4.3	3.4	—
Standard deviation ($\pm$)						
50 km	2.2	2.8	1.9	2.2	1.6	2.2
100 km	2.2	2.8	2.0	2.3	1.7	—
Ice types						
First year (%)						
<3.0 m						
50 km	40.6	49.8	72.9	27.2	57.4	54.7
100 km	46.0	48.4	60.2	25.3	54.1	—
Multi-year/deformed (%)						
>3.0 m						
50 km	59.4	50.2	27.1	73.8	43.6	45.3
100 km	54.0	51.7	39.8	74.7	45.9	—
Polynyas/leads						
Open water/new ice (%) <0.3 m						
50 km	0.0	0.8	2.5	0.1	0.0	0.1
100 km	0.0	1.6	1.4	0.1	0.0	—

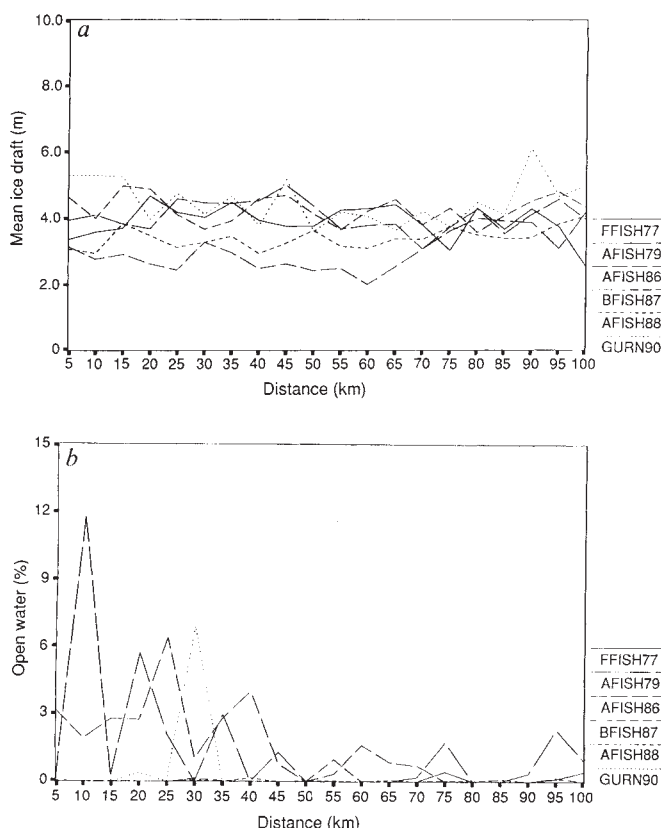


FIG. 1 a, Mean ice draft of 5-km segments during 100-km transect over the North Pole. b, Percentage of open water in these segments.

ice' and ice with draft greater than 3.0 m is assumed to be 'multiyear or deformed ice'. Additionally, any ice draft less than 0.3 m is assigned to the 'open water/new ice' category.

The results reveal considerable variation in all ice-draft categories during the six years. Especially noteworthy is the interannual variability revealed in the mean ice draft (± 1.0 m) and open water extent ($\pm 2.5\%$) as shown in Figs 1 and 2. The fractional coverage of multi-year ice also varies strongly on an interannual basis, for example from 27.1% in 1986 to 73.8% in 1987. These interannual variations are sufficient to preclude definitive conclusions about recent changes in the Arctic maritime environment. For example, the average of the ice drafts over 50-km segments from the late 1970s (1977, 1979) is 4.10 m, whereas the average over 50-km segments from the late 1980s (1986–88, 1990) is 3.45 m. The difference of 0.65 m indicates an apparent thinning of 15% from the late 1970s to the late 1980s (interestingly, the loss of ice volume north of Greenland between 1976 and 1987 was also 15%, according to Wadhams's⁵ analysis of British submarine data). But a *t*-test based on the means and variations of the subsamples from the 1970s ($N_1 = 2$) and 1980s ($N_2 = 4$) indicates that the difference of the subsample means is significant at no better than the 20% level ($t = 1.58$). In other words, the subsamples are so small and the interannual variations so large that there is a 20% probability that the subsample means will differ by chance by as much as 0.65 m. If the mean drafts of successive years are not independent because of the autocorrelation timescales of the ice-draft anomalies, then one may have even less confidence that the differences of the subsample means are meaningful. Application of similar tests to the other quantities in Table 1 produces even smaller (less significant) *t*-statistics. We thus conclude that the data do not provide evidence of a sustained trend towards change in sea-ice draft or type at the North Pole during the period from the late 1970s to 1990.

Statistical analysis of under-ice draft distribution data recorded by the six submarines throughout the central Arctic Basin ($> 89^\circ$ N) indicated similar results to those presented in Table 1: mean ice drafts and interannual variability were similar to that measured over the North Pole. This suggests that the variability observed over the North Pole is widespread and more

likely to be associated with interannual variability than with sampling variability.

To the extent that ice draft is proportional to ice thickness, our results also pertain to ice thickness. Our analysis has emphasized the interannual variability of various properties of sea ice near the North Pole, primarily because this profoundly affects the interpretation of apparent changes or trends. With regard to the detectability of a trend in the sonar-derived data, several points deserve mention. First, Gloersen and Campbell⁷ have recently reported a statistically significant decrease of total Arctic ice extent during 1978–1987, which is roughly the period spanned by our data. By contrast, we focus on ice draft in the central portion of the transpolar drift stream. Each of these variables has its own physical significance, and it is possible that one dataset might justify a statement about trends while the other does not. Second, our results are directly relevant to the ice thickness comparison (1976 compared with 1987) presented by Wadhams⁵, who found that the ice volume north of Greenland was ~15% less in 1987 than in 1976. The same decrease is apparent if one compares our draft profiles of 1986–1990 with those of 1977–1979. But the multi-year dataset analysed here permits us to extend Wadhams' results by showing that the interannual variations are large enough to prevent the assignment of statistical significance to the decrease. Given this variability, the probability of obtaining the same decrease by chance is at least 20%. Although this probability may be small enough to imply a meaningful trend to some users of scientific results, it is too high for many others. Additional measurements

over the next several years will clearly be valuable.

The results have other implications. The interannual variability of the open-water percentages from 0 to 3–4% points to a need in global climate models for lead and polynya parameterizations that are more sophisticated than the simple specification of a constant lead/polynya fraction. When air temperatures are ~–30 °C, ocean-atmosphere heat fluxes will be vastly different over pack ice sectors having lead/polynya fractions of 0 and 3%. Finally, the large interannual variability of the multi-year sea-ice fractions indicates either that the ice concentrations over a particular area at the end of the summer melt season are highly variable from year to year, or that the ice dynamics and deformation vary sufficiently to permit the formation, on a year-to-year basis, of very different amounts of new ice between freeze-up (September) and the following April. The implications for ocean-atmosphere heat exchange are again substantial. □

Received 29 April; accepted 2 June 1992.

1. Intergovernmental Panel on Climate Change *Climate Change: The IPCC Scientific Assessment* (eds Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) (Cambridge Univ. Press, 1990).
2. Manabe, S., Spelman, M. J. & Stouffer, R. J. *J. Clim.* **5**, 105–126 (1992).
3. Bourke, R. H. & Garrett, R. P. *Cold. Reg. Sci. Technol.* **13**, 259–280 (1987).
4. McLaren, A. S. *J. geophys. Res.* **94**, 4971–4983 (1989).
5. Wadhams, P. *Nature* **345**, 795–797 (1990).
6. McLaren, A. S. *et al. Nature* **345**, 762 (1990).
7. Gloersen, P. & Campbell, W. J. *Nature* **352**, 33–36 (1991).
8. Parkinson, C. L. & Cavalieri, D. J. *J. geophys. Res.* **94**, 14499–14523 (1989).
9. McLaren, A. S. *Arctic* **41**, 117–126 (1988).

ACKNOWLEDGEMENTS. This work was supported by a NSF Arctic Programs SGER grant.

Longitudinal confinement of geomagnetic reversal paths as a possible sedimentary artefact

C. G. Langereis*, A. A. M. van Hoof* & P. Rochette†

* Paleomagnetic Laboratory 'Fort Hoofddijk', Budapestlaan 17, 3584 CD Utrecht, The Netherlands

† Laboratoire de Géosciences de l'Environnement URA-CNRS 132, Faculté des Sciences Saint-Jérôme, Avenue Escadrille Normandie Niemen, 13397 Marseille cedex 13, France

THE positions of the virtual geomagnetic pole (VGP) during a large number of reversals of the Earth's magnetic field seem to show a remarkable confinement to longitudes over the Americas or to antipodal longitudes^{1,2}, although it has been argued that this confinement is statistically insufficiently constrained³. It has also been pointed out⁴ that the same bands of longitude appear in other geophysical observations, such as the pattern of fluid motion in the outer core and regions of higher seismic velocities in the lower mantle, suggesting a causal relationship. Here we show that longitudinal confinement of VGPs can arise from the smoothing of non-antipodal stable directions (just) before and after a geomagnetic reversal, because of the filtering effect of the remanence acquisition process in sediments. The origin of this non-antipodality is still uncertain and must remain speculative until more reversal records become available that include sufficiently long pre- and post-transitional intervals.

Many new palaeomagnetic reversal records have been reported in recent years, mostly from sedimentary sequences. These records have served as a test for reversal models. For instance, the low-order zonal harmonics model of Williams and Fuller⁵ predicts that VGP paths should be confined either to the site longitude (near sided) or to its antipode (far sided). Results^{1,2} suggesting a confinement independent of site longitude either to a path over (approximately) the Americas or to antipodal longitudes over Southeast Asia and Australia raised considerable excitement in the palaeomagnetic and geomagnetic community⁶,

mainly because of the suggested relationship with other geophysical observations⁴. But there is reason for considerable caution in interpreting palaeomagnetic records from relatively slowly deposited sediments. Evidence from volcanic records indicates different transitional field behaviour including wide longitudinal scatter (see for example refs 7–10), although it has been suggested that particular VGP clusters recur¹¹. More important, many aspects of the process of remanence acquisition and retention in sediments, such as authigenesis and (early) diagenesis, are still largely unknown. Recently it was shown that the palaeomagnetic components in two sedimentary reversal records have a delayed remanence acquisition and probably do not reflect the true geomagnetic field during a transition¹². In addition, it was pointed out earlier¹³ that smoothing due to remanence acquisition in a zone long with respect to reversal duration produces a record in which intensity changes take longer than directional changes, as is usually observed in sediments. This smoothing may further produce the typical confined VGP paths found in sedimentary records, particularly in the case of non-antipodal stable directions (just) before and after a reversal. Non-antipodality in declination typically yields VGP paths 90° W or 90° E of the observation site^{3,13,14}, whereas non-antipodality in inclination (inclination anomaly^{15–18}) leads to a near-sided or far-sided path⁵.

To test the effect of sediment smoothing, we have analysed reversals recorded in late Miocene marine marls from Crete¹⁴ and in the Pliocene Trubi marls from Calabria¹⁹ and Sicily (refs 12, 19, 20; and unpublished data). All records have typical sedimentation rates of 5 ± 1 cm kyr⁻¹. These records were chosen because the established magnetostratigraphies for Crete²¹, Calabria²² and Sicily²³ allow the determination of non-transitional stable directions. These are the mean directions of the entire underlying and overlying polarity zones with durations that range from ~60 to more than 400 kyr (refs 21–24); they explicitly exclude directions close (less than ~50 cm) to a transition. To determine whether there is a more pronounced non-antipodality of stable directions close to the actual transition, caused for example by geomagnetic instabilities that are related to the reversal^{25–27}, we have in addition determined the mean near-

transitional stable directions, where available and determined from the reversal records themselves. We have checked the (non-)antipodality of both non-transitional and near-transitional mean directions by using the reversal test²⁸ (Table 1), and found that in almost all cases these directions are significantly non-antipodal. We refer to the resulting filtered VGP paths (see Fig. 1) as the non-transitional model and the near-transitional model, respectively.

Caution must be used in interpreting near-transitional directions. Because of the general lack of reversal records that include sufficiently long pre- and post-transitional sampled intervals, some intervals used are less than typical post-depositional remanent magnetization (PDRM) lock-in depths^{29,30} (15–20 cm), and this may introduce an unwanted bias. But even if

a longer interval is used, authigenic magnetite formation may cause a chemical remanent magnetization (CRM) to be acquired down to depths of 80–120 cm (refs 12, 20). The substantial CRM acquisition depth may well bias the near-transitional directions as an average of stable polarity and (true?) transitional directions. In addition, this process may be complicated by the fact that CRM acquisition probably occurs only (or mainly) in specific depth intervals that depend on cyclically changing redox conditions²⁰, leading to a discontinuous smoothing filter much more complex than applied here.

For the Cretan records, it seems that the observed VGPs generally follow a path close to that determined by the non-transitional model (Fig. 1). The near-transitional model shows less agreement with the observed VGPs for the older two records (KS02 and KS05). In KS05, VGPs are initially close to the non-transitional model, but in a later stage seem to 'hesitate' between the latter and the near-transitional path. For the high-resolution records KS06 and especially KP₄01, however, the agreement of the near-transitional model with the observed VGPs is better. Furthermore, for the younger reversal, KP₄02, there is now considerably better agreement: the observed VGPs initially follow the non-transitional path over North America, but then 'jump' to the near-transitional path over Australia. This switching between antipodal VGP paths is a pattern of behaviour often seen in sedimentary records.

For the Calabrian records, the results of filtering generally agree well with the observed VGPs (Fig. 2). Although the lack of intermediate VGPs for the lower Sidufjall makes any comparison unreliable, the better resolution of the lower and upper Nunivak again provides a very good correlation between observed and filtered VGPs, either non-transitional or near-transitional. In the lower Nunivak, both observed and filtered VGPs show a near-sided path, whereas the upper Nunivak shows the 'more familiar' path over the Americas.

From the Sicilian records (Fig. 3), the lower Mammoth VGP path¹⁹ is poorly established, but the observed VGPs still agree with both models and pass over Southeast Asia. The non-transitional model predicts the same Asian path for the upper Mammoth¹⁹, but the observed path lies west of the Americas and is in good agreement with the near-transitional model. Our most recent and very-high-resolution records from Sicily are those of the Thvera and Nunivak subchrons (ref. 12; and unpublished data). Because of earlier observations of 'excursions' due to delayed remanence acquisition¹², care was taken to sample

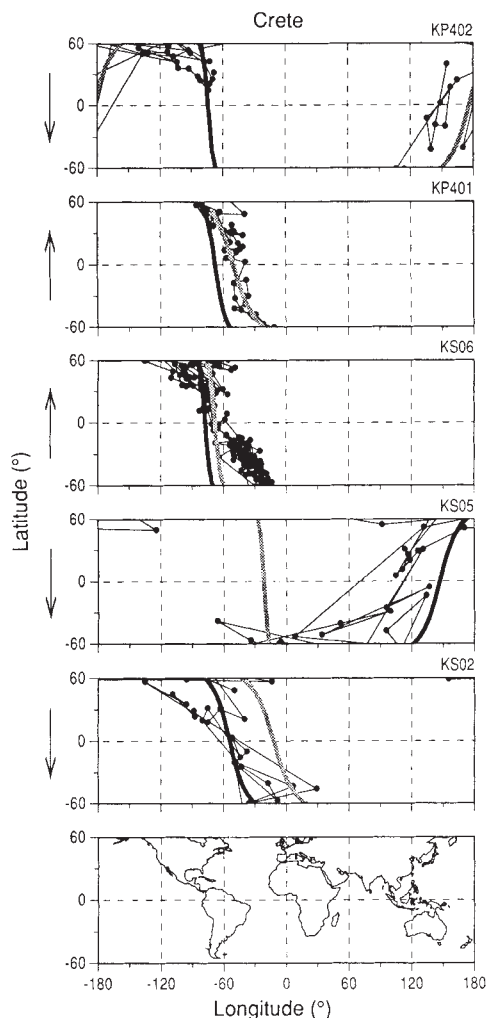


FIG. 1 Intermediate VGPs (those between latitudes 60° S and 60° N) of upper Miocene reversal records from Crete¹⁴. KS02 and KS05 are two older reversals, KS06 and KP₄01 represent the same reversal recorded in two different sections, KP₄02 is the youngest reversal (codes follow ref. 14). The arrow indicates the sense of the reversal; the lowermost frame presents the continents for reference. The solid line shows the VGP path obtained by filtering the mean directions of the entire under- or overlying polarity zones²¹ (non-transitional directions), the shaded line represents filtering of near-transitional directions as determined from the reversal records (see also Table 1). Filtering was done by assuming an instantaneous reversal from the pre-transitional to the post-transitional stable directions (both of unit intensity) using a moving rectangular window that 'sees' an increasing portion (0–100%) of the usually non-antipodal post-transitional direction; a triangular or gaussian window gives essentially the same results¹³. The resulting intermediate directions have been used to calculate the corresponding intermediate VGPs.

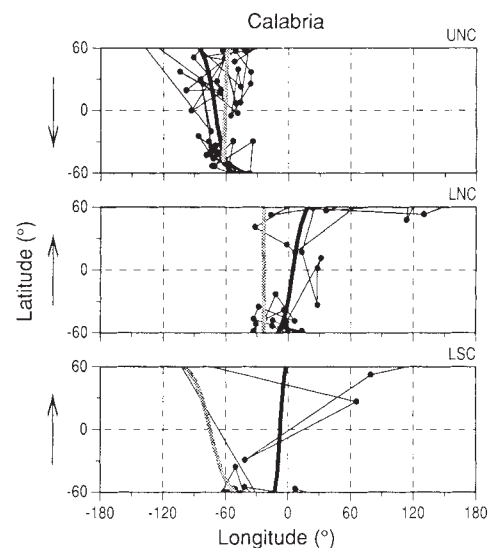


FIG. 2 Intermediate VGPs of lower Pliocene reversal records from Calabria¹⁹. LS, US indicate lower, upper Sidufjall; LN, UN represent lower, upper Nunivak. Post-fix C refers to Calabria. See also Fig. 1.

TABLE 1 Results of reversal test

Reversal	Sense		Interval (cm)		N1	R1	N2	R2	γ	γ^*	ctmd
			pre	post							
Crete											
KP ₁ 03	(R-N)	non	650	600	10	9.940	8	7.940	16.8	8.3	negative
		near	22	34	5	4.988	6	5.932	12.7	8.9	negative
KP ₁ 02	(N-R)	non	850	650	12	11.900	10	9.940	11.8	5.4	negative
		near	78	226	12	11.875	7	6.910	20.7	7.6	negative
KP ₄ 02	(N-R)	non	900	350	15	14.770	4	3.990	14.1	6.4	negative
		near	32	25	6	5.900	6	5.604	8.9	19.9	Class C
KP ₁ 01	(R-N)	non	1,900	850	28	27.750	12	11.900	15.8	4.8	negative
		near	87	91	25	24.810	10	9.854	29.7	6.3	negative
KP ₄ 01	(R-N)	non	3,000	900	22	21.770	15	14.770	14.3	5.6	negative
		near	39	67	10	9.838	11	10.775	33.3	9.1	negative
KS06	(R-N)	non	1,600	700	14	13.760	9	9.930	14.6	7.1	negative
		near	95	110	33	32.786	34	32.634	21.6	5.4	negative
KS05	(N-R)	non	300	1,600	5	4.970	14	13.760	2.3	7.7	Class B
		near	74	57	6	5.776	7	6.504	20.5	21.9	Indeterm.
KS02	(R-N)	non	400	700	7	6.962	8	7.980	9.8	5.3	negative
		near	26	17	11	10.769	11	9.950	10.4	16.6	Class C
Calabria											
UNC	(N-R)	non	1,000	1,000	29	28.811	16	15.128	10.4	9.4	negative
		near	8	6	19	18.261	19	18.814	37.3	7.5	negative
LNC	(R-N)	non	1,250	1,000	20	19.888	29	28.811	4.1	3.2	negative
		near	60	44	26	25.970	20	19.260	10.4	7.1	negative
USC	(N-R)	non	350	1,250	14	13.782	20	19.888	7.6	5.8	negative
		near	23	36	17	16.346	60	59.426	14.5	7.5	negative
LSC	(R-N)	non	550	350	3	2.991	14	13.782	7.6	5.8	negative
		near	43	35	16	15.387	73	71.594	13.6	8.2	negative
UTC	(N-R)	non	950	550	12	11.696	3	2.991	3.1	9.9	Class C
		near	27	30	30	28.074	52	51.380	13.9	7.5	negative
Sicily, upper Pliocene											
UKS	(R-N)	non	350	2,600	12	11.950	32	31.750	9.3	3.6	negative
		near	9	21	25	24.035	52	51.199	14.7	6.3	negative
LKS	(N-R)	non	350	350	8	7.860	12	11.950	9.2	8.2	negative
		near	26	11	15	14.674	22	20.975	5.6	9.3	Class B
UMS	(R-N)	non	450	350	9	8.922	8	7.860	8.8	9.2	Class B
		near	15	10	15	14.942	16	15.602	15.8	6.6	negative
LMS	(N-R)	non	1,150	450	19	18.923	9	8.922	7.0	5.7	negative
		near	15	15	28	27.783	23	22.418	11.7	5.4	negative
Sicily, lower Pliocene											
GGS	(R-N)	non	3,300	1,150	49	48.280	19	18.920	5.0	3.2	negative
		near	11	85	13	12.904	92	8.630	21.5	4.3	negative
UCS	(N-R)	non	450	3,300	8	7.950	49	48.280	12.3	5.9	negative
		near									
LCS	(R-N)	non	900	450	17	16.780	8	7.950	9.3	6.0	negative
		near									
UNS	(N-R)	non	650	900	12	11.930	17	16.780	10.1	5.2	negative
		near	8	25	37	35.381	94	92.511	37.3	5.6	negative
LNS	(R-N)	non	850	650	12	11.930	13	12.790	9.9	6.2	negative
		near	30	71	75	74.550	233	226.233	9.5	2.1	negative
LSS	(R-N)	non	500	400	7	6.930	4	3.980	12.5	8.6	negative
		near	15	52	11	10.748	27	26.185	27.6	8.6	negative
UTS	(R-N)	non	1,200	500	7	6.930	11	10.860	10.3	8.0	negative
		near	6	21	20	19.517	68	67.494	53.6	5.5	negative
LTS	(N-R)	non	500	1,200	11	10.860	4	3.960	7.8	10.7	Class C
		near	30	31	66	65.146	62	59.848	36.7	3.9	negative

Results of the reversal test²⁸ of both non-transitional (non) and near-transitional (near) stable mean directions, testing for a common true mean direction (ctmd) between the mean directions of two distributions (of N_1 and N_2 samples). The test is positive if the angle between the two means (γ) does not exceed the critical angle (γ^*); the latter depends on N_1 , N_2 and the respective unit vector sums $\mathbf{R}_1$, $\mathbf{R}_2$ (refs 28, 37) of the two distributions. The classification (A, B, C, indeterminate) depends on the value of γ^* , with breakpoint at 5°, 10° and 20°. The test is negative if $\gamma > \gamma^*$. The sense of the reversal (N-R or R-N) is given. Intervals used to determine near- and non-transitional directions before (pre) and after (post) the transition are given in centimetres. Most cases show a negative test, indicating statistically significant non-antipodality; only in a few cases is an indeterminate or a B/C classification found. Abbreviations are as in Figs 1–3; KP₁01, KP₄01 and KS06 represent the same reversal from different sections¹⁴, as do KP₁02 and KP₄02, and the same reversal may be sampled both in Calabria and Sicily (for example UTC and UTS). Near-transitional directions are taken as stable (that is, non-intermediate and approximately near-axial dipole field) directions from the actual reversal records, overlapping only slightly or not at all with the 'non-transitional' intervals. In some records, the near-transitional interval is of the order of typical PDRM lock-in depths^{29,30} and the filtering results may thus be biased.

long and detailed records, and near-transitional directions are accurately established. The agreement with the near-transitional model is particularly good.

The analysis of these central Mediterranean records shows that there is good agreement with the non-transitional model and an even better correlation with the near-transitional model. We have analysed in total 23 reversal records, 19 of which show intermediate VGPs in good or excellent agreement; one record is undetermined because of a lack of intermediate directions (LSC, Fig. 2), and three give VGPs that are antipodal with respect to the model. Hence, more than 80% of the records are successfully modelled by smoothing of non-antipodal directions. Furthermore, depending on the degree of smoothing, a clockwise swing in declination (an American VGP path for a reversed-normal (R-N) transition) may appear as a counterclockwise swing in the record³¹ (a Southeast Asian/Australian VGP path). This could provide a possible explanation why some records show VGPs antipodal to the expected path. In addition, a differential degree of smoothing may well explain why several records such as the upper Sidufjall in Calabria¹⁹, the lower Kaena¹⁹ and upper Kaena²⁰ from Sicily, show two antipodal bands of VGPs within one and the same transition. Such behaviour is not easily explained by models involving a (non-reversing) equatorial-symmetric dipole¹ or nondipole² com-

ponent, although it may be compatible with a reversing equatorial dipole¹¹. Our analysis raises several important questions: what is the extent of smoothing and/or why are non- and near-transitional directions not antipodal? But first we discuss the simplest possible cause, an unremoved overprint.

A combined analysis of the palaeomagnetic²³ and carbonate³² data from the Pliocene sediments on Sicily suggests that a higher carbonate content causes increased porosity, leading to increased weathering and thus to a larger secondary (normal) overprint. Rock magnetic studies showed that such an overprint may persist even to high (demagnetization) temperatures^{33,34}. A similar constant unremoved overprint was reported from the Deccan traps basalts, but this may be a bias from early palaeomagnetic data³⁵. It affects 'primary' (syndepositional or near-depositional) normal and reversed directions differently, depending on tectonic tilting and/or rotation. For instance, no tilting or rotation, plus an inclination error of the primary component, will lead to non-antipodality in inclination only, whereas superposition of a secondary overprint on a clockwise rotated primary component (35° in Sicily; 15° in Calabria) leads to non-antipodality in declination with a westerly offset and hence to VGPs over America. Although this mechanism may largely explain the (high-carbonate) lower Pliocene data, it does not explain the strong easterly non-antipodality of the (low-carbonate) upper Pliocene mean directions, with a clockwise rotation of 25–35°, that show no or only a small overprint³⁶. This suggests that at least here the observed non-antipodality may have a different origin.

Non-antipodality in inclination is well established^{15–18}, and sediment smoothing of this inclination anomaly may have been the cause of the many observed near- and far-sided VGP paths that prompted the Williams and Fuller model⁵. A declination anomaly is less well established¹⁶. From a recent analysis of outer core flux patterns²⁷, however, a qualitative model for the time-averaged geomagnetic field was derived that shows that none of the time-averaged normal declinations are significantly different from zero, but reversed declinations are, in some latitudinal regions in the Atlantic hemisphere. Filtering of these declinations (in our case defined as non-transitional) would lead to a predominantly westerly VGP path in the Atlantic hemisphere, in agreement with recent observations^{1,2,4} and with most of our own observations (Figs 1–3). Furthermore, even the significant CRM acquisition depths cannot explain the non-antipodality of polarity zone mean directions that represent the average of usually much longer intervals (3–30 m; Table 1).

The good fit of the near-transitional model may have a similar (non-antipodality) origin, but may also be related to the fact that it 'sees' a considerable part of the transition. Although the wide longitudinal scatter observed in volcanic records^{3,7–10} seems to provide no reason for the observed confinement, it has been argued that particular VGP clusters (in South America and Australia) may recur¹¹. Indeed, smoothing of such a temporal feature would bias the resulting magnetizations¹¹ so that they would often agree with our observations. Although similar VGP clusters (near South America) are seen in some of our 'continuous' sedimentary records (for example UNC, Fig. 2; UNS, Fig. 3), the complex and 'discontinuous' CRM filter prevents a firm assessment of their validity.

Apparently, the information that we may derive from sedimentary records (which are probably biased, possibly continuous, but often insufficiently long) and from volcanic records (discontinuous, possibly unbiased, but too few) is still far from complete³⁸. □

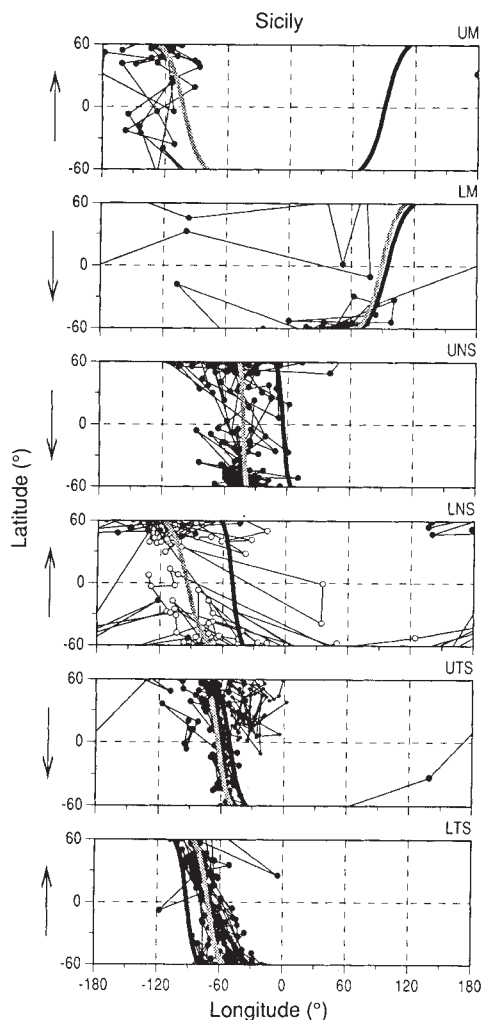


FIG. 3 Intermediate VGPs of Pliocene reversal records from Sicily (refs 19, 20 and unpublished data). LT, UT indicate lower, upper Thvera; LN, UN as in Fig. 2; LM, UM indicate lower, upper Mammoth. Post-fix S refers to Sicily. See also Fig. 1. Open symbols in LNS refer to the low-temperature component¹², as the high-temperature component (closed symbols) shows almost no intermediate directions.

Received 28 November 1991; accepted 4 May 1992.

1. Tric, E. *et al.* *Phys. Earth planet. Inter.* **65**, 319–336 (1991).
2. Clement, B. M. *Earth planet. Sci. Lett.* **104**, 48–58 (1991).
3. Valet, J. P., Turcholka, P., Courtillot, V. & Meynadier, L. *Nature* **356**, 400–404 (1992).
4. Laj, C., Mazaud, A., Weeks, R., Fuller, M. & Herrero-Bervera, E. *Nature* **351**, 447 (1991).
5. Williams, I. & Fuller, M. *J. geophys. Res.* **86**, 11657–11665 (1981).

6. Bogue, S. *Nature* **351**, 446 (1991).
7. Coe, R. & Prévot, M. *Earth planet. Sci. Lett.* **98**, 292–298 (1989).
8. Prévot, M., Markinen, E. A., Grommé, C. S. & Coe, R. S. *Nature* **316**, 230–234 (1985).
9. Chauvin, A., Roperch, P. & Duncan, R. A. *J. geophys. Res.* **95**, 2727–2752 (1990).
10. Roperch, P. & Duncan, R. A. *J. geophys. Res.* **95**, 2713–2726 (1990).
11. Hoffman, K. A. *Nature* **354**, 273–277 (1992).
12. van Hoof, A. A. M. & Langeris, C. G. *Nature* **351**, 223–225 (1991).
13. Rochette, P. *Earth planet. Sci. Lett.* **98**, 33–39 (1990).
14. Valet, J.-P., Laj, C. & Langeris, C. G. *J. geophys. Res.* **93**, 1131–1151 (1988).
15. Cox, A. *Rev. Geophys.* **13**, 35–51 (1975).
16. Merrill, R. T. & McElhinny, M. W. *The Earth's Magnetic Field* (Academic, London, 1983).
17. Schneider, D. A. & Kent, D. V. *J. geophys. Res.* **93**, 11621–11603 (1988).
18. Gubbins, D. *J. geophys. Res.* **93**, 3413–3420 (1988).
19. Linssen, J. H. *Geol. Ultratincta* **80**, 230 (1991).
20. van Hoof, A. A. M. & Langeris, C. G. *J. geophys. Res.* **97**, 6941–6958 (1992).
21. Langeris, C. G. *Geol. Ultratincta* **34**, 180 (1984).
22. Zijdeveld, J. D. A., Zachariasse, W. J., Verhallen, P. J. J. M. & Hilgen, F. J. *Newsl. Strat.* **16**, 169–181 (1986).
23. Langeris, C. G. & Hilgen, F. J. *Earth planet. Sci. Lett.* **104**, 211–225 (1991).
24. Berggren, W. A., Kent, D. V., Flynn, J. J. & Van Couvering, J. A. *Geol. Soc. Am. Bull.* **96**, 1407–1418 (1984).
25. Olsen, P. *Phys. Earth planet. Inter.* **33**, 260–274 (1983).
26. McFadden, P. L. & Merrill, R. T. *Phys. Earth planet. Inter.* **43**, 23–33 (1986).
27. Gubbins, D. *Nature* **326**, 167–169 (1987).
28. McFadden, P. L. & McElhinny, M. W. *Geophys. J. Int.* **103**, 725–729 (1990).
29. Valet, J. P., Tauxe, L. & Clement, B. M. *Earth planet. Sci. Lett.* **94**, 371–384 (1989).
30. deMenocal, P. B., Ruddiman, W. F. & Kent, D. V. *Earth planet. Sci. Lett.* **99**, 1–13 (1990).
31. Hoffmann, K. A. & Slade, S. B. *Geophys. Res. Lett.* **5**, 483–486 (1986).
32. Hilgen, F. J. & Langeris, C. G. *Terra Nova* **1**, 409–415 (1989).
33. van Velzen, A. & Zijdeveld, J. D. A. *Geophys. Res. Lett.* **17**, 791–794 (1990).
34. van Velzen, A. & Zijdeveld, J. D. A. *Geophys. J. Int.* (in the press).
35. Vandamme, D., Courtillot, V., Besse, J. & Montigny, R. *Rev. Geophys.* **29**, 159–190 (1991).
36. Zachariasse, W. J., Zijdeveld, J. D. A., Langeris, C. G., Hilgen, F. J. & Verhallen, P. J. J. M. *Mar. Micropalaeontol.* **14**, 339–355 (1989).
37. Fisher, R. A. *Proc. R. Soc. A* **217**, 295–305 (1953).
38. McFadden, P. *Nature* **356**, 381 (1992).

ACKNOWLEDGEMENTS. We thank V. Courtillot, M. J. Dekkers and J. D. A. Zijdeveld for comments on the initial manuscript, D. Gubbins and J. P. Valet for their constructive reviews, and P. J. Verplak for his help in the field and in the laboratory. The research was supported by the Netherlands Foundation for Earth Sciences (AWON) with financial aid from the Netherlands Organization for Scientific Research (NWO).

Link between geomagnetic reversal paths and secular variation of the field over the past 5 Myr

Catherine Constable

Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, La Jolla, California 92093-0225, USA

PALAEOMAGNETIC records provide information about the behaviour of the geomagnetic field during reversals^{1,2}. Existing records are incompatible with transitional field configurations that are either entirely dipolar or entirely zonal (dependent only on latitude)^{3,4}. Recent compilations^{5–8} have indicated that the transitional paths of virtual geomagnetic poles (VGPs) for the past few reversals are located preferentially within two antipodal longitudinal bands, suggesting that simple but non-zonal field configurations dominate during reversals. Here I point out that one of the longitudinal bands coincides with that expected from the reversal of a non-axial-dipole field exactly like that present today; the other requires only a sign change in the non-axial-dipole terms of today's field. Evidence for persistent non-zonal contributions to the field has generally^{9–13} (but not always^{14,15}) been regarded as not statistically significant in the light of poor data distributions. I show here that a non-zonal bias, similar to that observed in reversal data, is evident in data on secular variation of the field over the past 5 Myr, even after normalization according to site locations. These results suggest that the time-averaged field does indeed contain persistent (but not constant) non-zonal contributions.

Reversal records from a single location are usually plotted in the form of projections of VGP positions, with lines connecting adjacent time horizons. Each VGP position gives the point of intersection with Earth's surface of the axis of the geocentric

dipole that would generate the measured palaeomagnetic declination and inclination for a single time horizon at that site¹⁰. As the geomagnetic field polarity moves from a reversed to normal state, details of the migrations of VGPs from extreme southern to extreme northern latitudes provide a view of the reversal process. The advantage of the VGP transformation is that it removes the geographical dependence of the dipole part of the field: if the field elements of an exclusively dipolar field are measured at any location on the globe, they will all transform into a single VGP position; similarly, for a reversal with exclusively dipolar transitional fields, a single transition path will be obtained regardless of the location from which the observations are drawn (see Fig. 1). It is unlikely that the field is exclusively dipolar during reversals. As with the present-day field, one should expect non-dipole components to be an integral part of the field: such complexities are described by a spherical harmonic expansion, in which the geomagnetic field **B**, at radius *r*, colatitude *θ* and longitude *φ*, is written as the gradient of a harmonic potential, *Ψ*

$$\Psi(r, \theta, \phi) = a \sum_{l=1}^{\infty} \sum_{m=0}^l \left(\frac{a}{r}\right)^{l+1} (g_l^m \cos m\phi + h_l^m \sin m\phi) \times P_l^m(\cos \theta) \quad (1)$$

P_l^m are partially normalized Schmidt functions¹⁰ describing the latitudinal variation of the field. At any time the Gauss coefficients, *g_l^m* and *h_l^m*, determine the size of spatially varying components of the field that increase in complexity with spherical harmonic degree *l*. Zonal fields are those for which all Gauss coefficients with *m* ≠ 0 are zero; they have no longitudinal variations. For the present normal field the axial-dipole term *g₁⁰* is overwhelmingly dominant, and the next-largest contributions come from the non-zonal, equatorial dipole terms, *g₁¹* and *h₁¹*. Sampling the 1980 magnetic field at sites distributed worldwide yields VGPs clustered near the north geomagnetic pole, with a tilt away from the Earth's axis and a longitudinal bias reflecting the predominant influence of *g₁¹* and *h₁¹*. As for a dipole field, for an exclusively zonal field there is a simple relationship between VGP positions and the spherical harmonic representation; VGPs from any zonal field are constrained to lie along the great circle containing the site longitude. For example, a zonal quadrupole field superimposed on a geocentric axial dipole will generate VGP positions that lie on the far side of the geographic pole relative to the measurement site. But the presence of non-zonal fields (*m* ≠ 0 terms) complicates the issue, as the VGP transformation is nonlinear. In stable polarity times, persistent non-zonal terms may generate VGP positions with non-axisymmetric distributions about the geographic axis.

VGP paths for reversing dipole fields

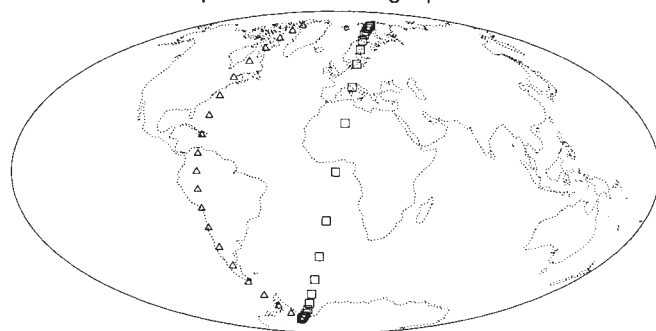


FIG. 1 Transitional VGP paths for reversals simulated from an exclusively dipolar field. Triangles are for reversal during which the equatorial contributions to the dipole remain static throughout and the VGP longitude remains fixed at 70.8° W; squares show a reversal in which the equatorial contributions reverse sign along with the axial part of the dipole, and the VGP longitude migrates from 70.8° W to 109.2° E.

During a reversal, when the axial-dipole part of the field decays in intensity, changes in the relative sizes of non-zonal contributions can considerably alter VGP locations.

A recent compilation^{5,6} of 21 records of the Matuyama-Brunhes geomagnetic reversal (dated at 0.78 Myr; ref. 16) showed that the VGP paths had some distinctive features. First, the individual records tended to produce paths confined to fairly narrow longitude bands. This phenomenon has often been observed in sedimentary records, and may originate from either geomagnetism or rock magnetism. If it is geomagnetic then the significant non-axial-dipole contributions to the field (all coefficients in the spherical harmonic expansion except g_1^0) are stable for the duration of the reversal; that is, they are much less variable than the axial dipole. Alternatively, the geomagnetic field variations may be smoothed during the acquisition of remanence (this process may continue for thousands of years after the sediment is deposited) and the result is then some average of the field configuration during and after the reversal¹⁷⁻¹⁹. Second, when the reversals are considered collectively and the distribution of all the transitional VGPs is plotted as a function of longitude, there are two roughly antipodal bands in which the bulk of the VGPs are concentrated (Fig. 2a). Clement⁶ modelled three of these reversal paths by a dominant h_3^1 term superimposed on a reversing axial dipole, and suggested that this indicated geographical control of the reversal path. He also noted an apparent correlation between favoured transition paths and concentrations of flux in recent radial field models at the core-mantle boundary²⁰; these flux concentrations may be reflected in full polarity palaeopole positions back to 5 Myr (ref. 14), but they do not appear in core-mantle boundary models unless terms of degree $l=4$ or higher are included. A selective compilation^{7,8} of records from reversals spanning the last 11 Myr showed similar confinement of VGP paths and suggested comparatively long time constants for the geographic control. It is possible, however, that these features reflect the geographical location of the sampling sites and inadequate normalization for the different numbers of transitional VGPs in each record²¹. How the number of VGPs in records from different sites should be normalized is not obvious, because the number of transitional VGPs at a given location depends on the predominant transitional field configuration, as well as whatever temporal variability in the sampling is enforced by the recording medium.

An interesting feature of the apparent control of reversal paths by core-mantle boundary conditions is that one of the regions of longitudinal confinement includes the position of the current geomagnetic pole. If the axial-dipole part of today's geomagnetic field decayed and grew back with the opposite polarity, while leaving the rest of the field (the non-axial-dipole part) as it is today, one would see one of the longitudinal bands shown by the reversal data. The longitudinal distribution of VGP data from such a simulated reversal, sampled at 21 equally spaced time points and 788 roughly uniformly distributed points on Earth's surface, is shown in Fig. 2b. Reversing the signs of the coefficients of the non-axial-dipole part of the field, and leaving those terms constant during the reversal as before generates an antipodal band. Combining the distributions generated by two such simulated reversals produces a figure like that obtained for the Matuyama-Brunhes transition (Fig. 2c). The effect of spatial distribution of reversal sites may be assessed in Fig. 2d, where the simulated reversals are sampled at the same geographical locations as the Matuyama-Brunhes data used by Clement⁶; the influence on the resulting VGP longitude distributions is small.

What does this similarity in VGP longitude distributions mean? The low geomagnetic field intensity during reversals makes rocks more susceptible to residual overprinting by the field present now, or by the field present immediately after the reversal. This could certainly account for the path through the Americas, but it is less obvious why there should be an antipodal band for data from a single reversal. The simplistic reversal

model used to generate Fig. 2b is not adequate either; the twin band at a single time instant requires dominant non-zonal spherical harmonic contributions of degree two or higher⁶. Alternatively, the bimodal distribution for reversal paths may result from inadequate temporal resolution in the sedimentary records, with each mode reflecting a preferred general configuration for the geomagnetic field. If transition between these states occurs sufficiently rapidly, VGPs in both longitude bands could arise during one reversal, and with adequate temporal resolution could be detected at a single site (as sometimes occurs). A global distribution of sites sampling over a long time period would generate VGP longitudes occurring over both the Americas and in the band over Western Australia and East Asia. The present field would provide a snapshot of one preferred configuration. Since AD 1600 the geomagnetic dipole location has migrated

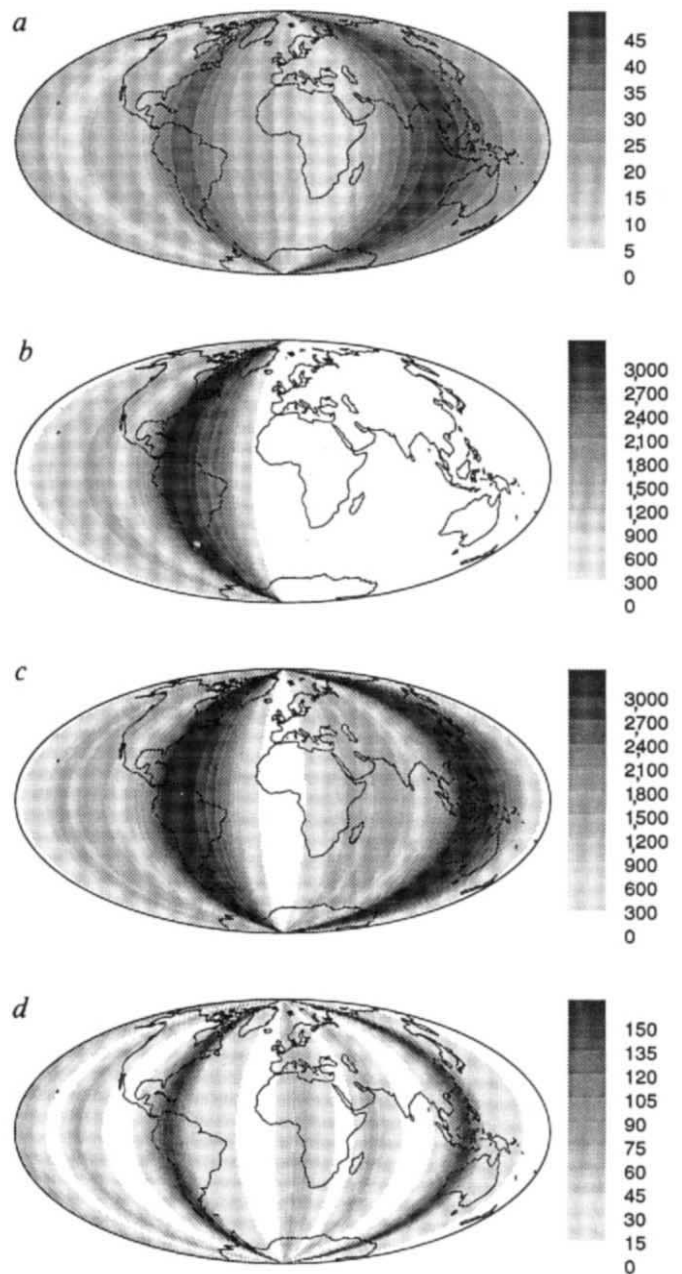


FIG. 2 Longitudinal distributions of VGP positions from (a) Matuyama-Brunhes reversal records, (b) a reversal simulated from the 1980 field sampled at 788 uniformly distributed sites, (c) a twin reversal simulated from the 1980 field distributed sites and (d) same as c but for those sites for which Matuyama-Brunhes records are available.

from roughly 40° W to 70° W (ref. 10); at 1600 it was still close to (but not exactly over) the peak VGP occurrence in Fig. 2b. Some reversal records provide evidence for long-lived preferred VGP locations²², a result that is compatible with the idea that the time-averaged field has a significant non-zonal component.

If such a bias exists, then one would also expect it in the average field configuration derived from palaeosecular variation (PSV) data. Figure 3a shows a histogram of the longitudinal distribution of VGP positions for PSV data for the past 5 Myr. Each position is an average result for several samples from a lava flow. There are 2,244 globally distributed data from 88 studies (see refs 23, 24). In Fig. 3a they have not been corrected for plate motions, because in many cases the reported ages are

not exact enough to provide accurate corrections over such a short time span. The effect of such motions was shown to be negligible by adjusting all the data sites to their positions 5 Myr ago and repeating the analysis. Using lava flow data minimizes the effects of time-averaging during remanence acquisition. Most studies of the time-averaged field⁸⁻¹⁵ have used palaeomagnetic pole positions¹⁰ (not VGPs) drawn from both sediments and basalts; these are averages of VGP directions, designed to remove the effects of secular variation. The most striking feature of the plot is that the PSV VGP distributions have a structure similar to that seen in the Matuyama-Brunhes reversal data and in the simulated twin reversal: they are most concentrated in two antipodal longitude bands. The regions are broader than those for the present-day field, or for the reversal, an effect that would be readily accounted for by noise in the palaeomagnetic data and a better temporal average of the secular variation. The longitudinal distribution of sites around the world is non-uniform: there are no data from sites between 99° W and 27.2° W or 77.5° E and 119.5° E (Fig. 3b). To compensate for this, I normalized contributions to the histogram of VGP longitudes so that equivalent weight was given to longitudinal regions of width 45°, except for the bin between 90° W and 45° W, which contains no data. The effects of this normalization are small (dashed line on Fig. 3a). Although the normalization cannot entirely compensate for the uneven data distribution, the lack of correlation between the VGP distribution of Fig. 3a and site distribution in Fig. 3b does argue strongly against a zonal field configuration. The VGP longitudes for the PSV data are obviously not well modelled by a uniform distribution; both Kuiper's V_N statistic and the Kolmogorov-Smirnov statistic D_N (ref. 25) give significance levels well below 1%.

The PSV data support the idea of preferred longitude bands in the VGP distributions. The details of this distribution are liable to change as more data becomes available, but it is clear that the zonal models mostly used in determining the average geomagnetic field configuration over the past 5 Myr are inadequate, and we need to look for models that show the minimum necessary departure from the geocentric axial dipole field, including longitudinally varying spatial components. The time constants for such non-zonal contributions have traditionally been thought to be short, because of westward drift velocities and also because at some sites palaeomagnetic directions become on average those of an axial dipole in only a few thousand years; but short time constants do not preclude the existence of a bias in the average field. The bias in VGP longitudes is unlikely to be seen in the standard checks^{10,13} for residual equatorial dipole contributions, g_1^1 and h_1^1 , because the pair of bands observed is not well modelled by an $l=1$ term. Such analyses have not been done for terms of higher degree. Global palaeosecular variation from lake sediments and archaeomagnetic sites could also be used to obtain better resolution than is possible with either reversal or PSV data. These span several thousand to tens of thousands of years and would provide a useful test of these ideas for prehistoric times. □

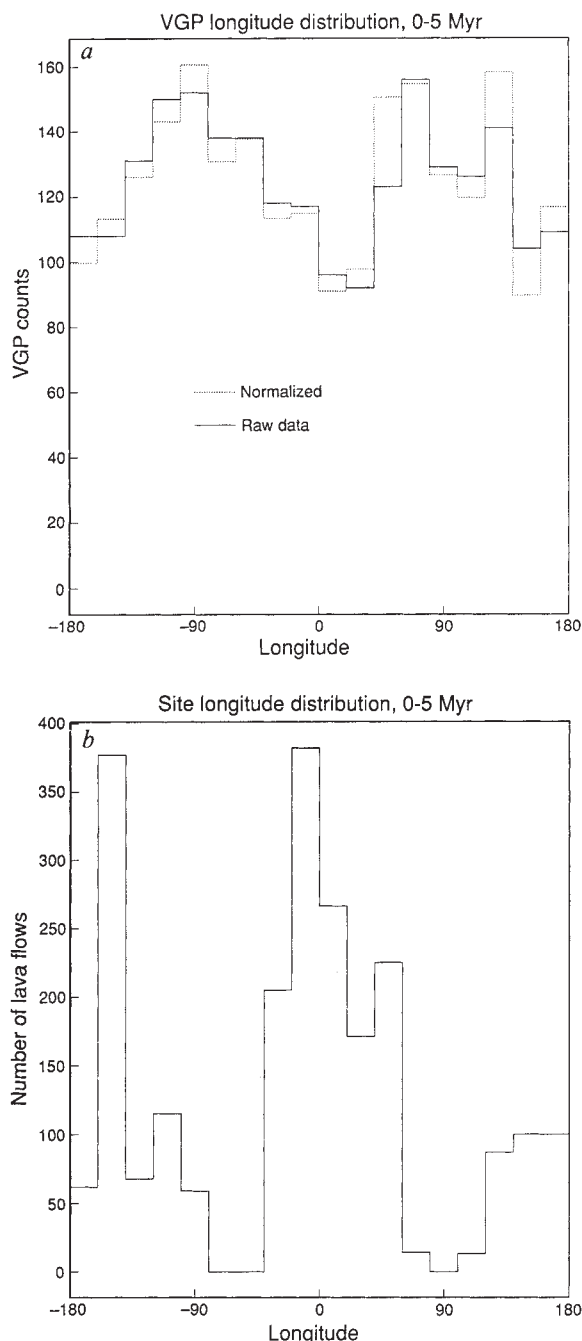


FIG. 3 a, Histogram of VGP longitudes for globally distributed palaeosecular variation data from lava flows spanning the past 5 Myr. Solid line is for the raw data, dashed line is normalized for site distributions. b, Histogram of site distributions as a function of longitude (20° bins) for the data in a.

Received 24 February; accepted 15 June 1992.

1. Bogue, S. W. & Hoffman, K. A. *Rev. Geophys.* **25**, 910-916 (1987).
2. Clement, B. M. & Constable, C. G. *Rev. Geophys. Suppl.: U.S. Nat. Rep. IUGG 1987-1990*, 433-442 (1991).
3. Hillhouse, J. & Cox, A. *Earth planet. Sci. Lett.* **29**, 51-64 (1976).
4. Hoffman, K. A. *Science* **196**, 1329-1332 (1977).
5. Clement, B. M. *EOS* **70**, 1073 (1989).
6. Clement, B. M. *Earth planet. Sci. Lett.* **104**, 48-58 (1991).
7. Tric, E. et al. *Phys. Earth planet. Inter.* **65**, 319-336 (1991).
8. Laj, C., Mazaud, A., Weeks, R., Fuller, M. & Herrero-Bervera, E. *Nature* **351**, 447 (1991).
9. Coupland, D. H. & Van der Voo, R. *J. geophys. Res.* **85**, 3529-3548 (1980).
10. Merrill, R. T. & McElhinny, M. W. *The Earth's Magnetic Field* (Academic, London, 1983).
11. Lee, S. & Lilley, F. E. M. *J. Geomagn. Geoelectr.* **38**, 797-806 (1986).
12. Merrill, R. T., McFadden, P. L. & McElhinny, M. W. *Phys. Earth planet. Inter.* **64**, 87-101 (1990).
13. Schneider, D. A. & Kent, D. V. *Rev. Geophys.* **28**, 71-96 (1990).
14. Gubbins, D. *J. geophys. Res.* **93**, 3413-3420 (1988).
15. Livermore, R. A., Vine, F. J. & Smith, A. G. *Geophys. J. R. astr. Soc.* **73**, 153-171 (1983).
16. Shackleton, N. J., Berger, A. & Peltier, W. R. *Trans. R. Soc. Edinb. Earth Sci.* **81**, 251-261 (1990).
17. Constable, C. G. *J. geophys. Res.* **95**, 4587-4596 (1990).

18. Rochette, P. *Earth planet. Sci. Lett.* **98**, 33-39 (1990).
19. Weeks, R. J., Fuller, M. & Williams, I. *J. geophys. Res.* **93**, 11613-11620 (1988).
20. Bloxham, J. & Gubbins, D. *Nature* **325**, 509-511 (1987).
21. Valet, J. P., Tucholka, P., Courtillot, V. & Meynadier, L. *Nature* **356**, 400-407 (1992).
22. Hoffman, K. A. *Nature* **354**, 273-277 (1991).
23. Lee, S. thesis, Australian National Univ. (1983).
24. McFadden, P. L., Merrill, R. T., McElhinny, M. W. & Lee, S. *J. geophys. Res.* **96**, 3923-3934 (1991).
25. Fisher, N. I., Lewis, T. & Embleton, B. J. *J. Statistical Analysis of Spherical Data* (Cambridge Univ. Press, 1987).

ACKNOWLEDGEMENTS. I thank C. Johnson and R. Parker for discussions, and the NSF for financial support.

Post-Jurassic mammal-like reptile from the Palaeocene

Richard C. Fox^{*†}, Gordon P. Youzwysyn^{*}
& David W. Krause[‡]

^{*} Laboratory for Vertebrate Paleontology, Departments of Geology and Zoology, University of Alberta, Edmonton, Alberta T6G 2E9, Canada
[‡] Department of Anatomical Sciences, State University of New York, Stony Brook, New York 11794-8081, USA

[†] To whom correspondence should be addressed

MAMMAL-LIKE reptiles of the Order Therapsida document the emergence of mammals from more primitive synapsids¹ and are of unique zoological and palaeontological interest on that account². Therapsids, first appearing in the Early Permian³, were thought to become extinct in the Middle Jurassic^{4,5}, soon after the Late Triassic origin of mammals⁶. Here, however, we report the dis-

covery of a therapsid from the late Palaeocene, 100 million years younger⁷ than the youngest previous occurrence of the order. This discovery nearly doubles the stratigraphic range of therapsids and furnishes their first record from the Cenozoic. The documenting fossils, an incomplete dentary containing three teeth, and four isolated teeth from other, conspecific individuals (Fig. 1), are from the Paskapoo Formation, at Cochrane, Alberta, Canada, from beds yielding a diverse mammalian fauna of early Tiffanian age⁸. These specimens are catalogued in the collections of the University of Alberta Laboratory for Vertebrate Paleontology (UALVP) and provide the basis for a new taxon, as named and described below.

Class Reptilia

Subclass Synapsida

Order Therapsida

Suborder Cynodontia

Chronoperatidae fam. nov.

Type genus *Chronoperates* gen. nov.

Chronoperates paradoxus gen. et sp. nov.

Etymology. *Chronos* (Greek): time; *perates* (Greek): wanderer, in reference to the geochronologic gap between this taxon and other known therapsids; *paradoxus* (Latin): contrary to expectations.

Holotype. UALVP 32358, left dentary having the three posteriormost teeth, two empty alveoli preceding these, and an incomplete coronoid process.

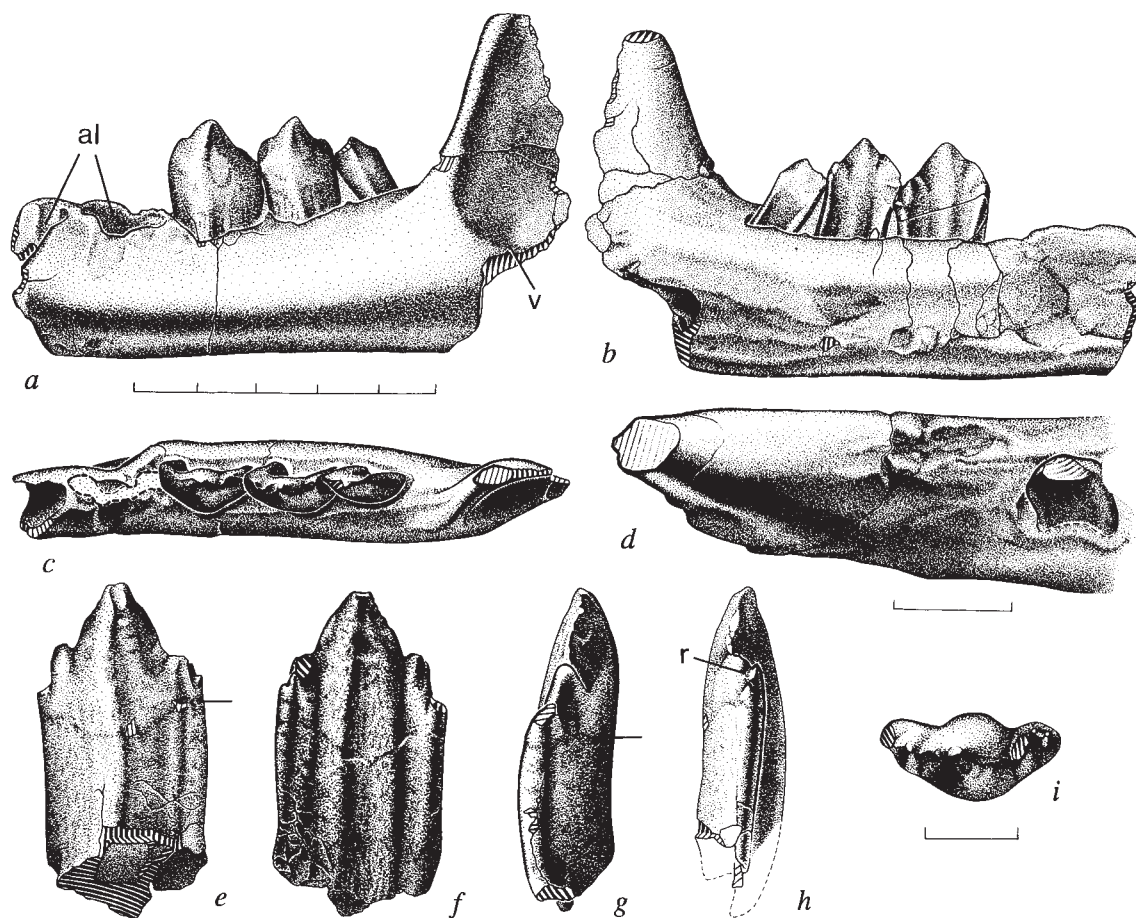
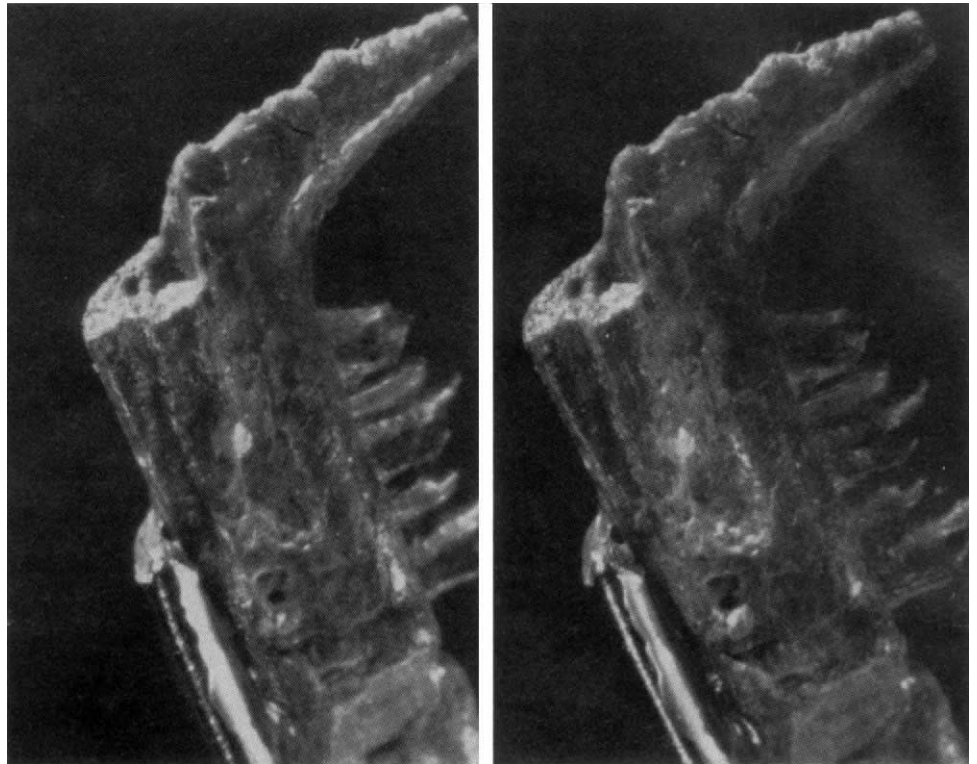


FIG. 1 UALVP 32358 (holotype), incomplete left dentary, *Chronoperates paradoxus* gen. et sp. nov., Cochrane 2 locality⁸, late Palaeocene, Alberta: a, labial view; b, lingual view; c, occlusal view; d, dorsal view of scar for coronoid bone behind posteriormost tooth. UALVP 32359, isolated lower left molariform tooth (distalmost extent of root missing): e, labial; f, lingual; g,

anterior; h, posterior; and i, occlusal views. al, Alveoli; v, ventral rim of masseteric fossa; r, posterior recess for receipt of next, more posterior tooth. Horizontal bar in e and g points to irregular line marking ventral extent of enamel. Scale bar divisions, 1 mm (a-c drawn to same scale; e-i drawn to same scale).

FIG. 2 UALVP 32358, showing posterior trough and adjacent surfaces of articulation for post-dentary bones (stereo pair, upper panel). h, Hook-shaped depression continuous with trough; t, trough for post-dentary bones; p, pit; s, scar for splenial and prearticular. Scale bar, 1 mm.



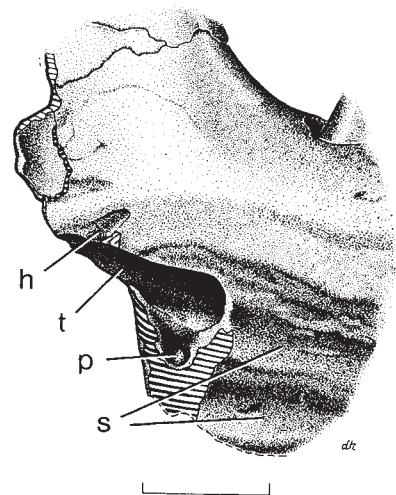
Type locality and stratigraphy. Cochrane 2, Alberta⁸; Paskapoo (= Porcupine Hills) Formation, late Palaeocene (early Tiffanian)⁸.

Type species. *Chronoperates paradoxus* sp. nov.

Diagnosis. This taxon differs from all others in its possession of tall, transversely narrow, single-rooted lower postcanine molariform teeth having three major cusps that form a broadly obtuse angle to a nearly straight line depending on tooth position; the teeth lack basal cingula, and the enamel has a pseudoprismatic microstructure.

The dentary (UALVP 32358; Fig. 1) is broken anteriorly and terminates posteriorly in a vertical coronoid process, which is incomplete. Laterally on the process is a masseteric fossa for insertion of external mandibular adductor muscles; this fossa is relatively small, with its ventral rim near the level of the alveolar row, well above the inferior border of the dentary (Fig. 1a). Posteromedially, near the base of the process and along the inner face of the dentary beneath the tooth row, are a trough and adjacent surfaces of articulation for one or more post-dentary bones, including the splenial (Fig. 2). A shallow Meckelian groove may have been developed, but crushing and breakage make a convincing demonstration of its presence impossible. No articular condyle is preserved, but from the dimensions of bone in the area, if one were originally present, it was small; a small dentary condyle is already present in the premammalian *Pachygenelus* and *Diarthrognathus*⁹.

The teeth are thecodont, each having a single elongate root and an undivided pulp cavity. The three posteriormost teeth are in place in UALVP 32358 and decrease in size posteriorly (Fig. 1a-c); X-rays indicate that the roots are fully formed, without bifurcation distally. More anteriorly are two empty alveoli that are broken and incomplete (Fig. 1a) but that also held single-rooted teeth, each roughly the size of the antepenultimate tooth on the specimen. Consequently, at least the five posteriormost lower postcanines were single-rooted. No replacing teeth or pits are evident externally, but a narrow bevelled surface runs lingually adjacent to the tooth row, in the position of the dental



lamina depression in some therapsids and primitive mammals; X-rays have failed to reveal replacing teeth in the dentary.

Overall, the teeth (Fig. 1) are unusually tall and very narrow transversely, although the crowns and roots of the teeth in UALVP 32358 are crescent-shaped, convex buccally. The crowns are basically tricuspid, with the central cusp the tallest. The cusp pattern, ranging from a broadly obtuse angle to a nearly straight line, evidently depends on tooth position: on UALVP 32358, the angle formed by the cusps of the anteriormost tooth is the most acute (Fig. 1c). On well preserved teeth, two small cusps arise posteriorly, where they interlock with the anterior edge of the next posterior tooth, but no talonid is developed; anteriorly on these teeth an additional cusp is present as well (Fig. 1e-i). The enamel grades smoothly on to the root without a basal cingulum or inflation of the crown beyond the dimensions of the root (Fig. 1e-i).

Chronoperates paradoxus is readily identified as a member of the synapsid plus mammalian clade¹ on the basis of the coronoid

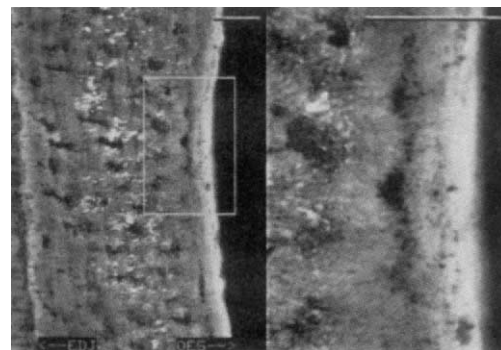


FIG. 3 Scanning electron micrograph of longitudinal section of enamel illustrating pseudoprismatic structure (that is, discontinuous but without prism boundaries). Area in rectangle outlined with white lines on left enlarged on right to show detail of pseudoprismatic structure deep-to-thin layer of aprismatic enamel. EDJ, enamel-dentine junction; OES, outer enamel surface. Scale bars, 10 μ m.

process, masseteric fossa and thecodont dentition¹⁰. Some advanced Diapsida (such as Lacertilia) possess a raised coronoid process formed by the coronoid bone or by the dentary and coronoid together¹¹⁻¹³, but the jaw adductors that are the analogues of the masseter in these species have an insertion on the coronoid, not on the dentary, and the dentition is pleurodont or acrodon^{12,13}. Primitively thecodont amniotes include archosauromorphs in addition to synapsids and mammals¹, but although some archosauromorph orders (Crocodylia, Pterosauria) contain species having single-rooted, multiple-cusped, 'mammal-like' teeth^{14,15}, these, like other archosauromorphs, retain the diapsid pattern of jaw structure and adductor muscle insertion^{14,15}. Clearly, the resemblances between *C. paradoxus* and all diapsids having a raised coronoid process and mammal-like dentition are convergent.

Our conclusion that *C. paradoxus* is a synapsid reptile¹, not a mammal, stems from both dental and mandibular characters. From overall resemblance of tooth crowns, a preliminary focus of our analysis, only three mammalian orders are relevant for comparison: Mesonychia, Edentata and Symmetrodonta. The first two can readily be dismissed: they are eutherian orders, and in them the mandibular skeleton is furnished by the dentary bone alone, so that a posteromedial trough and adjacent surfaces of articulation for post-dentary elements as seen on UALVP 32358 are lacking altogether; furthermore, the tooth enamel of *C. paradoxus* has a pseudoprismatic microstructure (Fig. 3), distinctly more primitive than the fully prismatic enamel seen in edentates (Simpson¹⁶ observed 'straight prisms' extending through the enamel in the Eocene armadillo *Utaetus*) and other eutherians¹⁷. As for symmetrodonts, which are pretribosphenic and the earliest therians, only the primitive *Kuehneotherium* is known to have a compound lower jaw^{18,19}. Symmetrodont lower molars are tricuspid, with the three cusps forming a triangular biting surface^{18,19}, but, as in all early mammals^{9,10}, symmetrodont lower postcanines are double-rooted¹⁸. If *C. paradoxus* is a symmetrodont, one must account

for retention of a primitive multiple-cusped, symmetrodont-like architecture of the molar crowns during evolution from a double-rooted to a highly derived single-rooted condition. We know of no mammalian examples in which the crowns of the postcanines remain unmodified during evolutionary reduction of the number of roots. From these considerations, we conclude that *C. paradoxus* is only convergent on the dentition of symmetrodonts, and note corroborating anatomical details: in the Cochrane species (1) the interlock between successive teeth is the reverse of that of symmetrodonts (and all other primitive mammals); (2) there is no basal cingulum; (3) the enamel lacks well defined shearing surfaces and prisms; and (4) in no known symmetrodont does the angle formed by the lower molar cusps become more obtuse posteriorly. In overall morphology the teeth of *C. paradoxus* most closely resemble those of certain Triassic cynodonts, including *Therioherpeton*²⁰, *Microconodon*²¹, and unnamed species from western Europe^{22,23}.

Among tetrapods, only some cynodonts and *C. paradoxus* combine: (1) single-rooted lower postcanines having transversely narrow, multiple-cusped crowns lacking cingula; (2) pseudoprismatic enamel; (3) retention of post-dentary bones, including a well developed splenial (cynodont lower jaws are illustrated in ref. 1); (4) small masseteric fossa. Hence, *C. paradoxus* must have branched from a cynodont clade that had not yet evolved the more derived, 'mammalian', states of these same characters, surviving then as a relict to the Palaeocene. In this context, the Cochrane fossils indicate that therapsids and mammals were contemporaries for at least the first two-thirds of mammalian history. The failure to discover therapsids in Late Jurassic through Middle Palaeocene strata may be attributable to one or more of several factors, including the following: (1) their rarity; (2) their extremely small size; (3) their possible presence in only more northerly regions; (4) limited sampling of the terrestrial vertebrate microfauna by use of underwater fine-screening techniques, particularly of Middle Jurassic through latest Cretaceous strata. □

Received 10 March; accepted 27 May 1992.

1. Carroll, R. L. *Vertebrate Paleontology and Evolution* (Freeman, New York, 1988).
2. Hopson, J. A. *Am. Biol. Teacher* **49**, 16-26 (1987).
3. Laurin, M. & Reisz, R. R. *Nature* **345**, 249-250 (1990).
4. Charlesworth, E. *Rep. Br. Ass. Adv. Sci.* **25**, 80 (1854).
5. Cui, G. & Sun, A. *Vert. Palaeont.* **25**, 245-259 (1987).
6. Fraser, N. C. *et al. Nature* **314**, 161-163 (1985).
7. Harland, W. B. *et al. A Geological Time Scale* (Cambridge Univ. Press, 1989).
8. Fox, R. C. in *Dawn of the Age of Mammals in the Northern Part of the Rocky Mountain Interior, North America* (eds Bown, T. M. & Rose, K. D.) 51-70 (Geological Society of America, Boulder, Colorado, 1990).
9. Shubin, N. *et al. Science* **251**, 1063-1065 (1991).
10. Hopson, J. A. & Barghusen, H. R. in *The Ecology and Biology of Mammal-like Reptiles* (eds Hotton, N., MacLean, P. D., Roth, J. J. & Roth, E. C.) 83-106 (Smithsonian Institution Press, Washington, DC, 1986).
11. Gould, S. J. *Earth Sci. Rev.* **6**, 77-119 (1970).
12. Gauthier, J., Estes, R. & de Queiroz, K. in *Phylogenetic Relationships of the Lizard Families* (eds Estes, R. & Pregill, G.) 16-118 (Stanford Univ. Press, Stanford, 1988).
13. Estes, R., de Queiroz, K. & Gauthier, J. in *Phylogenetic Relationships of the Lizard Families* (eds Estes, R. & Pregill, G.) 119-281 (Stanford Univ. Press, Stanford, 1988).

14. Clark, J. M. *et al. Science* **244**, 1064-1066 (1989).
15. Wild, R. *Boll. Soc. Paleont. Ital.* **17**, 176-256 (1978).
16. Simpson, G. G. *Bull. Am. Mus. nat. hist.* **91**, 1-232 (1948).
17. Carlson, S. J. in *Skeletal Biomineralization: Patterns, Processes and Evolutionary Trends* vol. 1 (ed. Carter, J. G.) 531-556 (Van Nostrand Reinhold, New York, 1990).
18. Prothero, D. R. *Bull. Am. Mus. nat. hist.* **167**, 281-325 (1981).
19. Cassiliano, M. L. & Clemens, W. A. in *Mesozoic Mammals, The First Two-thirds of Mammalian History* (eds Lillegraven, J. A., Kielan-Jaworowska, Z. & Clemens, W. A.) 150-161 (Univ. California Press, Berkeley, 1979).
20. Bonaparte, J. F. & Barberena, M. C. *J. Paleont.* **49**, 931-936 (1975).
21. Simpson, G. G. *Am. J. Sci.* **12**, 87-108 (1926).
22. Peyer, B. *Schweiz. Palaont. Abh.* **72**, 1-72 (1956).
23. Clemens, W. A. *Zitteliana* **5**, 51-92 (1980).

ACKNOWLEDGEMENTS. We thank D. Hollingdale for drawings, A. Lindoe and Gao K. for photography, A. Voss for sorting screened concentrates, G. Stonley for field assistance, F. Grine for critically reading the manuscript, and E. Dumont, F. Grine and L. Martin for assistance with preparation and examination of specimens with the scanning electron microscope. This work was partly supported by the Natural Sciences and Engineering Research Council of Canada, the NSF and the University of Alberta.

Targeted misexpression of *Hox-4.6* in the avian limb bud causes apparent homeotic transformations

Bruce A. Morgan*, Juan-Carlos Izpisua-Belmonte†‡, Denis Duboule† & Clifford J. Tabin*

* Department of Genetics, Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115, USA

† European Molecular Biology Laboratory, 6900 Heidelberg, Germany

‡ Present address: Department of Biological Chemistry, University of California, Los Angeles, California 90024-1737, USA

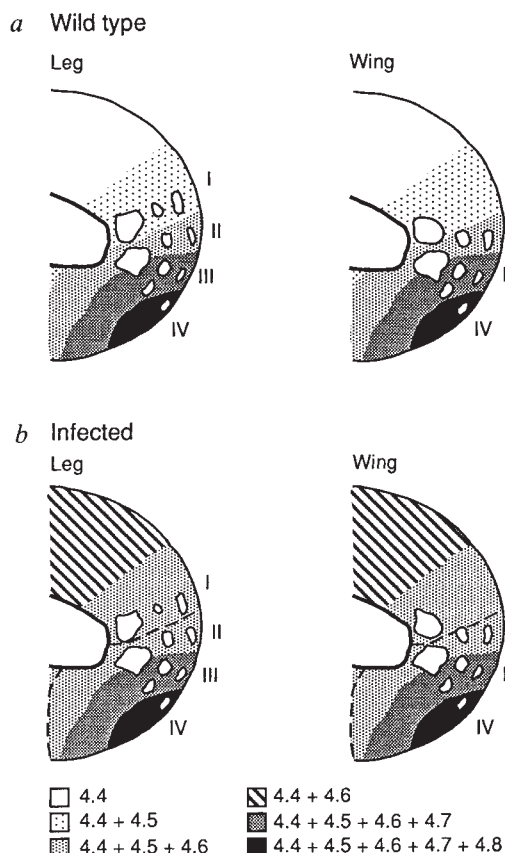
IN the limb bud the 5' members of the *Hox-4* gene cluster are expressed in a nested set of overlapping domains which are progressively restricted in the posterior and distal directions¹. These domains arise early in limb bud development and come to approximate the primordia of the major structural elements of the limb along the anterior/posterior axis² (Fig. 1). This pattern, and the fact that surgical manipulations which lead to mirror image duplications along the anterior/posterior axis give rise to mirror image duplications of the domains of expression of these genes, have led to the proposal that these transcription factors specify positional identity along the anterior/posterior axis^{3,4}. Here we test this hypothesis directly using replication-competent retroviral vectors to expand the domain of expression of the *Hox-4.6* gene anteriorly during limb development *in vivo*. We report that alteration of the domain of expression of the *Hox-4.6* gene in the developing limb leads to reproducible pattern alterations consistent with a posterior homeotic transformation.

The RCAS type-A replication-competent avian retroviral vector⁵ was used to alter *Hox-4.6* gene expression *in vivo*. The murine *Hox-4.6* complementary DNA was inserted into this vector, which retains all essential viral sequences and is capable of spreading from cell to cell in the infected chick embryo (Fig. 2a). The encoded murine protein shows 81% conservation (71% identity) to the chick *Hox-4.6* protein^{6,7}. Most of the divergence between these proteins is limited to two short blocks of poor conservation; the N terminus, central region and homeodomain are 90, 94 and 100% conserved, respectively. When transduced by the virus, the mouse gene is efficiently expressed in chicken cells (Fig. 2b).

Protocols were developed to achieve extensive infection of the limb bud and concomitant overexpression of the *Hox-4.6* gene throughout the bud during the period when limb mesenchymal cells are still developmentally plastic with respect to their ultimate positional identity along the anterior/posterior (A/P) axis (Fig. 2c). By injecting virus into the primordia of the limb at stage 10–12 (ref. 8), complete infection of the limb can be achieved by stage 21, when the precursors of distal structures can still respond to signals that alter their fate along the A/P axis⁹. Because of the kinetics of viral spread, infection and concomitant misexpression of the *Hox-4.6* gene are largely limited to the limb bud at this stage, although the infection will continue to spread as development proceeds.

Preliminary *in situ* analysis suggests that overexpression of the mouse *Hox-4.6* gene throughout the limb has no effect on the expression of the endogenous chick *Hox-4.4*, *4.6*, *4.7* or *4.8* genes (Fig. 3; and data not shown). If misexpression of the *Hox-4.6* gene has no effects on the other endogenous members of the cluster either, then the domains of *Hox-4* gene expression in a completely infected limb will resemble those shown in Fig. 1b. The domain defined by the concomitant expression of the *Hox-4.4*, *4.5* and *4.6* genes, which normally encompasses the

FIG. 1 Expression of the *Hox-4* genes in chick limb buds. **a**, Normal expression of the *Hox-4* genes in the stage-21 limb bud superimposed on a rough fate map of the bud (after ref. 11, and data not shown). Diagrammed in the upper panel are the domains of expression of the *Hox-4* genes (refs 2–4; and J.C.B., D.D. and B.A.M., data not shown) in a stage-22 leg (left) and wing (right) bud⁸. The *Hox-4* gene expression patterns change rapidly during development and the precise boundaries of expression cannot be rigorously aligned with a fate map at this early developmental stage before overt differentiation of structural elements has begun. Hence these diagrams serve only as approximations of the relative position of the *Hox-4* expression domains and the presumptive skeletal elements at this stage. The primordia of the four digits of the foot correspond roughly to the different domains of *Hox-4* gene expression in the bud at this developmental stage. Later in development, pre-cartilaginous condensations provide landmarks to align the fate map with *Hox-4* expression domains. At the stage shown (21), *Hox-4.6* expression is found in the digit II primordia and regions posterior to it whereas *Hox-4.7* is restricted to areas posterior to the anlage of digit II. As development proceeds, *Hox-4.7* is also expressed in more anterior regions (including the primordia of digit II). In a similar way, *Hox-4.8* is initially expressed in more posterior regions but spreads towards the anterior later in development. In contrast to the leg, the mature wing has only three digits (II, III and IV). As in the leg, the domain expressing *Hox-4.4* + *4.5* + *4.6* appears to encompass the primordia of digit II. But digit II is the most anterior digit of the mature wing; the region immediately anterior to the digit II primordia, which expresses only *Hox-4.4* and *4.5*, normally undergoes programmed cell death. **b**, Altered expression of the *Hox-4* genes produced by complete infection of the limb bud with the *Hox-4.6* virus. In the lower panel, the domains of expression of the *Hox-4* genes are diagrammed in a stage-22 leg (left) and wing (right) bud after complete infection with the *Hox-4.6* RCAS virus, assuming that expression of the mouse *Hox-4.6* protein does not effect the expression of the endogenous genes. The domain expressing *Hox-4.4* + *4.5* + *4.6*, which is normally restricted to post-axial tissue at this stage, is extended preaxially to encompass the precursors of digit I in the leg and a portion of the region of programmed cell death in the wing. Anterior to this region, a novel domain expressing *Hox-4.4* and *4.6* without *4.5* is created in both buds. In more posterior regions, the combination of *Hox-4* genes expressed is qualitatively unaffected.



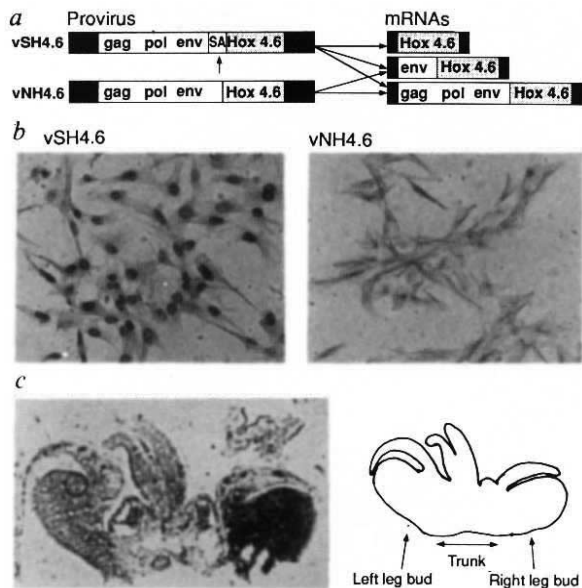


FIG. 2 The replication competent *Hox-4.6* virus. **a**, Structure of the *Hox-4.6* viruses. Viruses used in these experiments are diagrammed at left. The mouse *Hox-4.6* cDNA was inserted in the *Clal* site of the RCAS and RCAN replication-competent avian retroviral vectors derived from Rous sarcoma virus⁵. Both vectors retain the *gag*, *pol* and *env* genes required for the propagation of the virus. The RCAN vector lacks the splice acceptor (SA; arrow) found upstream of the insertion site in RCAS which is required for the expression of the inserted gene. Transcription begins in the 5' terminus of the provirus and the mRNAs are expressed by alternative splicing. The RCAS virus gives rise to the three major mRNAs diagrammed at right which include an mRNA for the *Hox-4.6* protein. The RCAN virus produces two principal mRNAs; the *Hox-4.6* mRNA is not produced by this virus. This virus serves as a control for nonspecific effects of viral infection and/or overexpression of *Hox-4.6* RNAs that are not mediated by the encoded protein. These constructs were transfected into chick embryo fibroblasts (CEF cells) and viral supernatants were collected 4 to 10 days later, after infection had spread through the culture.¹² **b**, Individual CEF cells infected with the *Hox-4.6* RCAS virus (left) express similar amounts of the *Hox-4.6* protein, cells infected with the *Hox-4.6* RCAN control virus (right) do not express the *Hox-4.6* protein. Infected cells were fixed in 4% paraformaldehyde and stained with a rabbit polyclonal antiserum raised against the mouse *Hox-4.6* protein (D.D., unpublished) followed by detection with a Vectastain alkaline phosphatase secondary detection kit (Vector Research). The *Hox-4.6* RCAS-infected cells show strong nuclear staining for the *Hox-4.6* transcription factor. Cells infected with the control virus (right) show only diffuse cytoplasmic staining indistinguishable from that observed with uninfected cells (data not shown). **c**, The entire limb bud is infected by stage 22 after microinjection at stage 11. Immunohistochemical staining of a horizontal section through the leg bud region of a stage-22 embryo which was infected *in ovo* shows complete infection of the injected limb bud. The dark alkaline phosphatase precipitation product indicates that most of the cells in the injected limb are infected. At this stage, the contralateral limb, as well as the rest of the embryo, is uninfected; alterations in gene expression are restricted to the injected limb. (To maintain the left-right orientation of the limb buds, the section is shown with dorsal-side down.) In an average injection run, 80–90% of the injected embryos showed complete infection of the limb bud by this stage. The remaining 10–20% showed partial infection at this stage. The embryo was windowed at stage 11, injected in the right leg primordia with the *Hox-4.6* virus as described¹³, and incubated until stage 22. The embryo was then killed, fixed overnight in 4% paraformaldehyde, dehydrated in 30% sucrose, embedded in gelatin and cryosectioned horizontally at the level of the leg buds. Sections were stained with the 3C2 monoclonal antibody against a Rous sarcoma virus p19^{gag} epitope¹⁴ followed by detection with a Vectastain alkaline phosphatase secondary detection kit (Vector Research).

primordia of digit II and posterior structures, is expanded in an anterior direction to encompass the primordia of digit I. In more posterior regions, the combination of *Hox* genes expressed is qualitatively unaffected. If the *Hox-4* genes specify fate along the A/P axis, expression of *Hox-4.6* throughout the bud should lead to skeletal transformations only in regions where the combination of *Hox* genes expressed has been altered. The developmental fate of cells in the primordia of the anterior digit I should be converted to that of the most posterior digit II. The fate of cells in more posterior regions of the bud should be unaffected.

Consistent with this model, in roughly 30% of the infected embryos the structure and number of the skeletal elements of digit I resembles those of digit II (Fig. 4a–d). An additional 10% of the infected individuals showed a partial conversion towards a posterior fate in the anterior skeletal elements of the foot. These results directly demonstrate that the *Hox-4* genes can control pattern along the A/P axis of the leg. Although here we concentrate on the effects on the clearly identifiable skeletal elements of the foot (see legend to Fig. 4), we observed other effects. These included reduction of the distal leg segment (zeugopod) and aberrations in foot morphology, including a planar arrangement of all four toes and suppression of interdigital cell death (Table 1).

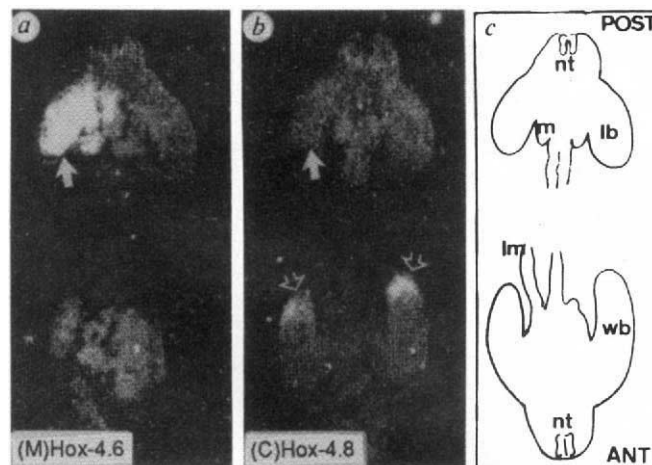
These other commonly observed phenotypes are also consistent with a posterior transformation of cell fate. The abbreviated distal leg segment may represent a partial conversion of the anteriorly derived tibia to the posterior fate of the shorter fibula. During limb outgrowth there is a progressive proximal-to-

distal loss of developmental plasticity and the failure, in some cases, to achieve complete conversion of the skeletal structures may indicate that total infection of the region and concomitant alteration of *Hox-4* gene expression was not achieved until after some developmental decisions had been completed irreversibly in the proximal elements of these limbs. Consistent with this explanation, the most frequently observed phenotypes involve alterations of relatively late morphogenetic events. Digit one is formed in the plane of the other digits and normally rotates to assume its final position opposing the other digits late in development. A planar foot in the absence of skeletal transformations may reflect a later conversion of digit I to a digit II identity which would therefore remain in the plane of digits III and IV. Interdigital cell death also occurs later in development after expression of the *Hox-4* genes has been restricted to the perichondrial regions and is no longer apparent in the interdigital regions. Persistent virally mediated expression of the *Hox-4.6* protein in the interdigital region apparently prevents the execution of the cell death program and directs these cells towards an alternative fate.

The differential penetrance of these phenotypes may reflect the relative probability of infecting every relevant cell of the limb bud before some distinct critical period for each developmental decision. In this context it is notable that delaying the onset of *Hox-4.6* misexpression by infecting the limb bud 12 hours later in development does not lead to any morphological transformations.

The observation that the *Hox-4* genes can control pattern

FIG. 3 Misexpression of exogenous *Hox-4.6* does not affect endogenous *Hox-4* gene expression. Adjacent sections from the dorsal side of an infected embryo were hybridized *in situ* to antisense RNA from the mouse (M) *Hox-4.6* and chicken (C) *Hox-4.8* genes^{4,14}. *a*, The exogenous *Hox-4.6* transcript is expressed throughout the injected leg bud (white arrow) but is not observed in the contralateral leg bud or wing buds. *b*, This section passes through the distal and posterior region of the wing buds where symmetrical domains of *Hox-4.8* expression are detected (hollow arrows). But this section passes through a more proximal/anterior plane of the leg buds where the endogenous *Hox-4.8* is not expressed in the control (contralateral) bud. *Hox-4.8* is also not detected in the infected (white arrow) limb bud. This experiment demonstrates that the exogenous *Hox-4.6* protein does not induce ectopic expression of the endogenous *Hox-4.8* gene. Similar preliminary experiments indicate that the expression domains of the endogenous *Hox-4.4*, *4.6* and *4.8* genes are unaltered by misexpression of the exogenous *Hox-4.6* protein. Endogenous *Hox-4.5* expression has not been examined. *In situ* hybridizations were done as described¹. *c*, Diagram of the plane of section. This section passes at an oblique angle through the trunk of the embryo at the level of the wing (wb) and leg (lb) buds and is viewed from the dorsal surface. The intervening trunk curves out of the plane of section. The dorsally located neural tube (nt) is seen at the extreme top and bottom of the section; ventral and lateral mesoderm (m and lm) are also labelled.

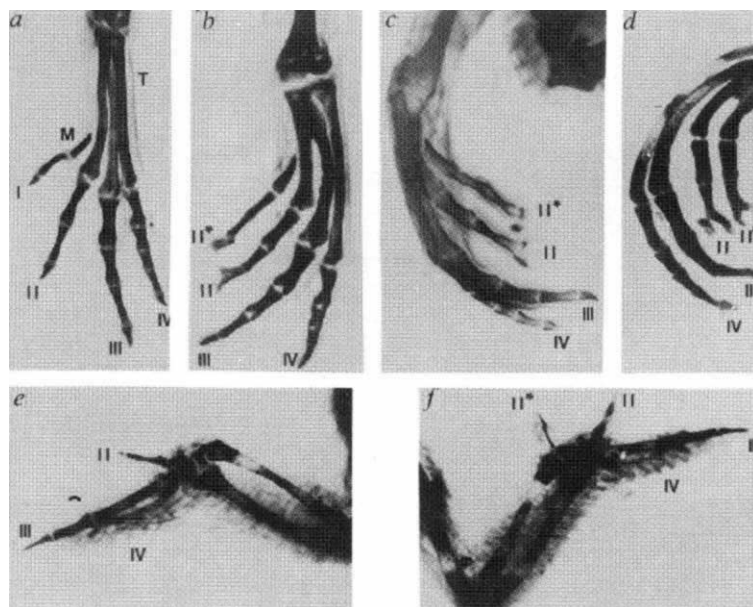


along the A/P axis of the limb was confirmed and extended by overexpression of *Hox-4.6* throughout the wing bud. There are three digits in the wing, designated II, III and IV (Fig. 4e). The anterior limit of the region of the limb bud that expresses *Hox-4.4* + *4.5* + *4.6* approximates the anterior edge of the primordia of digit II, as well as that of the ulna. Unlike the leg, however, the region immediately anterior to the digit II primordium that expresses *Hox-4.4* and *4.5* exclusively does not normally form a digit but undergoes programmed cell death. In wing buds infected with the *Hox-4.6* virus, an additional digit is frequently observed anterior to the normal digit II (46%; $n = 57$ injected individuals). In the best developed cases, this digit has a structure similar to that of digit II (Fig. 4f). These observations are

consistent with the hypothesis that the combination of *Hox-4* genes expressed in a given region specifies positional identity along the A/P axis and that expansion of the domain of expression of *Hox-4.6* in the anterior direction has converted these cells to a more posterior fate appropriate to the primordia of digit II.

All of the phenotypic effects observed in both the wing and the leg were specific to the misexpression of the *Hox-4.6* protein. Phenotypes were observed exclusively in the infected leg, the contralateral leg was unaffected. Control viruses (Fig. 2a) that do not splice the viral RNA to a functional *Hox* messenger RNA had no morphological effect ($n = 51$, leg; $n = 46$, wing). The pattern changes observed are not due to the nonspecific

FIG. 4 Homeotic conversions caused by ectopic expression of the *Hox-4.6* gene. *a*, The skeleton of a wild-type leg was stained with alcian blue and cleared through graded glycerols as described¹⁰ (photographed from the ventral side). The three posterior tarsometatarsals (T) arise at the tibiotarsal joint while the anterior metatarsal arises more distally (M). Excluding the terminal claw, the number of bones in a digit reflects its designation, digit I has 1 phalange, digit II has 2, digit III has 3, and digit IV has 4. *b*, An infected leg from an embryo that was injected in the right leg primordia with the *Hox-4.6* virus shows a partial transformation of the anterior digit (photographed from the ventral side). The anterior digit (II*) has two phalanges exclusive of the claw, although these bones are only partially separated at the joint. The anterior metatarsal is morphologically normal as are the bones of digits III and IV. All four digits remain in a plane and interdigital cell death has been suppressed between the anterior three digits (II*, II, III). *c*, An infected leg shows a partial transformation of the anterior digit (photographed from the dorsal side). The anterior digit (II*) has two phalanges exclusive of the claw, although these bones are only partially separated at the joint. The anterior metatarsal is morphologically normal as are the bones of digits III and IV. Interdigital cell death has been partially suppressed between digits II* and II and some adventitious cartilage is observed. (Suppression of interdigital cell death by other means can lead to the formation of similar adventitious cartilage nodules¹⁵.) *d*, An infected leg that exhibits a homeotic transformation of the anterior digit (II*) (photographed from the dorsal side). This digit shows the two bones (exclusive of the claw) which are characteristic of a digit II identity. The anterior metatarsal arises proximally at the tibiotarsal joint and resembles the tarsometatarsal which adjoins digit II. This specimen exhibits all of the phenotypes scored in Table 1, including suppression of interdigital cell death and the associated adventitious cartilage which gives rise to the forked appearance of the claws of digits II* and II, planar arrangement of all four digits, and abbreviated distal leg segment (not



shown). *e*, The left wing of an embryo injected in the right wing primordia is morphologically normal, exhibiting normal digits II, III and IV. *f*, The contralateral, injected limb bud from this embryo has a digit anterior to digit II. The bone structure and gross morphology of the additional digit (II*) resembles that of digit II. Embryos were infected and processed as described in the legend to Table 1 and photographed with a Wild stereomicroscope using transmitted illumination.

TABLE 1 Hox4.6 virus-infected leg phenotype summary ($n=100$ affected individuals)

Digit I to digit II†	30%
Digit I to digit III or IV†	0%
Digit II, III or IV to another fate‡	1%
Posterior conversion anterior digit or metatarsal§	40%
Abbreviated distal leg segments	52%
Planar foot	68%
Incomplete digital separation¶	93%

White leghorn embryos (SPAFAS, USA) were incubated until stage 11–12, injected *in ovo* in the right limb primordia with the *Hox-4.6* virus. The precise injection site was determined empirically to be the lateral to the neural tube ~0.5 mm from the posterior edge of the embryo. Eggs were allowed to develop until embryonic day 11. Embryos were killed, fixed overnight in 4% paraformaldehyde, and scored for obvious morphological defects. They were then dehydrated in ethanol, stained for cartilage and cleared as described¹⁰ and scored for skeletal alterations under a Wild stereomicroscope with transmitted illumination. A total of 131 individuals were scored in this study. Immunostaining of infected siblings at stage 21 indicated that in an average injection run 80–90% of the injected embryos showed complete infection of the limb bud by this stage. As 10–20% of the injected individuals are not adequately infected, only embryos that showed some morphological defect were included in the population analysed in this table. An additional 31 individuals showed no obvious phenotype.

† Digit identity was assigned by the number of phalanges in the digit. Cartilage segments containing incompletely formed joints were scored as separate phalanges if the contour of the nascent bone was consistent with this conclusion. 17% showed complete separation at the joint (Fig. 4d); 13% showed partial joints (Fig. 4b,c).

‡ In 100 injected animals that showed obvious pattern defects, one showed an apparent conversion of digit II to a digit III structure. 10 of these animals had severely malformed feet with partial truncation of digits IV or III. Although the reduced number of bones in these digits could be formally considered as anterior transformation, the grossly malformed appearance of these digits suggests that a less specific morphogenic defect is responsible. These individuals were not scored as interpretable transformations.

§ The anterior metatarsal giving rise to digit I arises distally and is quite short relative to the posterior tarsometatarsals. Elongated anterior bones which arose at the tibiotarsal joint were scored as partial conversions to a posterior fate.

|| Distal leg segment was less than 80% of the length of that of the contralateral leg.

¶ Partial (at least 1/3 of digit length) to complete webbing of one or more digits.

effects of overexpression of a homeobox transcription factor; the *Hox-4.4* gene is normally expressed throughout the limb and initial experiments suggest that retrovirus mediated overexpression of *Hox-4.4* in the limb bud has no phenotypic consequences (data not shown). Moreover, preliminary *in situ* hybridization analysis suggests that the ectopically expressed *Hox* mRNAs are not present at levels significantly above those derived from the endogenous genes (data not shown). The complexity of the phenotypes observed may be due in part to the temporal or quantitative alterations in *Hox-4.6* gene expression, or to dominant effects of the virally expressed genes on the normal targets of other *Hox* genes that are expressed in the limb. It should be possible to elucidate the roles different *Hox* genes play in limb development using the virally mediated misexpression approach.

Despite the complexity of the phenotypes, digit I was the only digit that routinely exhibited an apparent homeotic conversion; the more posterior digits which arise from regions that normally express the endogenous *Hox-4.6* gene were rarely converted to an inappropriate morphology (1%). Furthermore, the only alteration observed in digit I was that towards a digit II fate. The observation that alteration of the expression pattern of *Hox-4.6* can lead to an apparent homeotic conversion of digit I into digit II leads us to conclude that the *Hox-4* genes specify fate along the anterior/posterior axis of the developing limb.

Note added in proof: A revised vertebrate *Hox* gene nomenclature system has been provisionally adopted at the 1992 EMBL Homeobox Workshop. Under this scheme *Hox-4.4*, *-4.5*, *-4.6*, *-4.7* and *-4.8* are renamed *Hoxd-9*, *-10*, *-11*, *-12* and *-13*, respectively. □

Received 2 March; accepted 29 May 1992.

1. Dolle, P., Izpisua-Belmonte, J. C., Falkenstein, H., Renucci, A. & Duboule, D. *Nature* **342**, 767–772 (1989).
2. Yokouchi, Y., Sasaki, H. & Kuroiwa, A. *Nature* **353**, 443–445 (1991).
3. Nohno, T. *et al. Cell* **64**, 1197–1205 (1991).
4. Izpisua-Belmonte, J. C., Tickle, C., Dolle, D., Wolpert, L. & Duboule, D. *Nature* **350**, 585–589 (1991).
5. Hughes, S., Greenhouse, J., Petropoulos, C. & Sutcliffe, P. *J. Virol.* **61**, 3004–3012 (1987).
6. Izpisua-Belmonte, J. C., Falkenstein, H., Dolle, P., Renucci, A. & Duboule, D. *EMBO J.* **10**, 2279–2289 (1991).
7. Rogina, B., Coelho, C. N., Koshier, R. A. & Upholt, W. B. *Dev. Dynamics* **193**, 92–101 (1992).
8. Hamburger, V. & Hamilton, H. L. *J. exp. Morphol.* **88**, 49–92 (1959).
9. Summerbell, D. *J. Embryol. exp. Morphol.* **32**, 227–237 (1974).
10. Macleod, M. *J. Teratol.* **22**, 299–301 (1980).
11. Bowen, J., Hinchliffe, J., Horder, T. J. & Reeve, A. M. F. *Anat. Embryol.* **179**, 269–283 (1989).
12. Cepko, C. *Neuromethods* **16**, 177–219 (1989).
13. Dolberg, D. & Bissel, M. *Nature* **309**, 552–558 (1984).
14. Potts, W., Olsen, M., Boettiger, D. & Vogt, V. *J. gen. Virol.* **68**, 3177–3182 (1987).
15. Hurlie, J. M., Ganan, Y. *Anat. Embryol.* **176**, 393–399 (1987).

ACKNOWLEDGMENTS. We thank D. Fekete for working out the immunohistochemical procedures that allowed us to track the course of infection, and for technical advice on the use of retroviruses in the chick embryo; H. Falkenstein for technical assistance; and D. Fekete and R. Riddle for comments on the manuscript. B.A.M. was supported by an NRSA fellowship. This work was funded in part by a Public Health Service grant to C.J.T.

The Lowe's oculocerebrorenal syndrome gene encodes a protein highly homologous to inositol polyphosphate-5-phosphatase

Olivier Attree^{*†}, Isabelle M. Olivos^{*}, Ichiro Okabe^{†‡}, L. Charles Bailey^{*}, David L. Nelson[§], Richard A. Lewis^{§||}, Roderick R. McInnes[¶] & Robert L. Nussbaum^{*‡#}

^{*} Department of Human Genetics, and [‡] Howard Hughes Medical Institute, University of Pennsylvania School of Medicine, 422 Curie Boulevard, Philadelphia, Pennsylvania 19104–6145, USA

[§] Institute for Molecular Genetics and ^{||} Department of Ophthalmology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

[¶] Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada

LOWE'S oculocerebrorenal syndrome^{1–3} (OCRL) is a human X-linked developmental disorder of unknown pathogenesis^{4–8} and has a pleiotropic phenotype affecting the lens, brain and kidneys. The OCRL locus has been mapped to Xq25–q26 by linkage^{9–11} and by finding *de novo* X;autosome translocations at Xq25–q26 in two unrelated females with OCRL^{12,13}. Here we use yeast artificial chromosomes with inserts that span the X chromosomal breakpoint from a female OCRL patient in order to isolate complementary DNAs for a gene that is interrupted by the translocation. We show that the transcript is absent in both female OCRL patients with X;autosome translocations and that it is absent or abnormally sized in 9 of 13 unrelated male OCRL patients with no detectable genomic rearrangement. The open reading frame encodes a new protein with 71% similarity to human inositol polyphosphate-5-phosphatase. Our results suggest that OCRL may be an inborn error of inositol phosphate metabolism.

We assumed that the X;autosome translocations in females with OCRL disrupted the OCRL locus and therefore used yeast artificial chromosomes (yRS88 and yRS41) that span the

[†] Present addresses: CRSSA, Unité de Biologie Moléculaire, 38702 La Tronche Cedex, Grenoble, France (O.A.); and Department of Pediatrics, Jichi Medical School, Minamikawachi-Machi, Kawachi-Gun, Tochigi-Ken, Japan 329–04 (I.O.).

[#] To whom correspondence should be addressed.

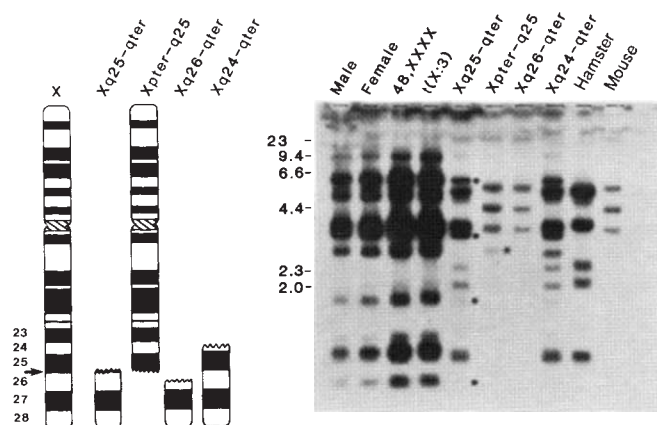


FIG. 1 Southern blot analysis of normal and hybrid cell line DNA shows that OCRL-1 cDNA overlaps the X chromosomal breakpoint in the female OCRL patient with a t(X;3) translocation. *a*, Schematic diagram of the regions of the human X chromosome retained in each of the four somatic cell hybrids in *b*. *b*, lane 1, normal male; lane 2, normal female; lane 3, 48,XXXX; lane 4, the t(X;3) female OCRL patient; lane 5, the human-hamster cell hybrid F649-5 which contains the der3 chromosome (Xq25-qter) from the t(X;3) female OCRL patient but lacks the derX and the normal X (unpublished data); lane 6, human-mouse cell hybrid Lo1-6TG containing the derX chromosome from same patient (Xpter-q25) (ref. 11) without the derivative 3 or normal X; lane 7, RJK734, a human-mouse cell hybrid retaining Xq26-qter²⁹; lane 8, human-hamster cell hybrid X3000-11 containing the region Xq24-qter as its only human component³⁰; lane 9, Chinese hamster V79 cell line; lane 10, mouse A9 cell line. Dots indicate fragments of human origin distal to the breakpoint; the asterisk indicates genomic fragments proximal to the breakpoint in somatic cell hybrid cells. Size calibration standards are shown on the left (in kb).

METHODS. Each genomic DNA (10 µg) was digested with *Bgl*II (5U of enzyme per µg of DNA) followed by extraction with phenol and chloroform, precipitation with ethanol and redigestion for several hours to completion. Restriction fragments were separated by electrophoresis, transferred onto Hybond N (Amersham), hybridized and washed according to the instructions of the manufacturer and standard protocols³¹. The probe was a 1,740-bp *Eco*RI fragment from the 5' end of λ HK7 insert, labelled by random priming³¹.

breakpoint in one of the female OCRL patients¹⁴ as probes to search cDNA libraries^{15,16}. Two partially overlapping clones containing ~4,350 base pairs (bp) of sequence (λ HK7 containing bases 1-2,320 and λ HK0 containing bases 1,625-4,353) were isolated from a human kidney cDNA library and shown to be derived from a gene that spans the t(X;3) breakpoint in this female OCRL patient. In the Southern blot shown in Fig. 1, a probe consisting of 1,740 bp at the 5' end of λ HK7

hybridized to many *Bgl*II fragments in DNA from normal male, normal female, 48,XXXX, and the t(X;3) patient. The human/hamster hybrid containing the der3 chromosome (Xq25-qter) from the t(X;3) patient (lane 5) revealed human fragments of 5.8, 3.2, 1.6 and 0.7 kilobases (kb) (Fig. 1), which must be distal to the X;3 translocation. In contrast, the human/mouse hybrid containing the derX (Xpter-q25) from the t(X;3) patient (lane 6) gave a 2.5-kb human *Bgl*II fragment (asterisk in Fig. 1) which must be proximal to the X;3 translocation breakpoint. Other human fragments could not be resolved from rodent fragments in the somatic cell hybrids. All human fragments detected by the 5' probe appear to map to the region Xq24-q26 as they are present in hybrid X3000-11 (Xq24-qter) (Fig. 1, lane 8) but absent in RJK734 (Xq26-qter) (Fig. 1, lane 7), as do all fragments detected by the full-length cDNA (data not shown).

Northern analysis of human tissue RNA using λ HK7 as probe revealed a single ~5.8-kb transcript present in brain, skeletal muscle, heart, kidney, lung and placenta, as well as in fibroblasts from three unrelated normal individuals, but not in immortalized lymphoblastoid cells from normals (data not shown). Northern analysis (Fig. 2) of RNA from fibroblasts from both female OCRL patients with X;autosome translocations and in 15 male OCRL patients (from 13 unrelated families) with no cytogenetic abnormality revealed no detectable transcript in both female translocation patients and in 8 unrelated male OCRL patients. One pair of affected brothers (B.B. and J.B.) lacked the normal 5.8-kb band and had an increased steady-state level of an abnormally large ~6.5-kb transcript. Another pair of affected brothers (W.J. and G.J.) and three other unrelated male patients had transcripts indistinguishable from controls in size and amount. All 9 unrelated male OCRL patients with abnormal or absent RNA transcripts had normal Southern blot patterns (data not shown); these results indicate that the messenger RNA abnormalities seen in patients were probably due to small alterations in this gene (such as a point mutation, or a small insertion or deletion).

The transcript (OCRL-1) represented by the consensus cDNA sequence (Fig. 3) of clones λ HK7 and λ HK0 has an open reading frame defined by an in-frame ATG triplet at base 16, just 9 bases downstream from a single in-frame stop codon at nucleotide position 7, and an in-frame stop at position 2,920, and is capable of encoding a polypeptide of 968 amino acids with a predicted M_r of 112,000 (112K). The transcript size, however, exceeds that of the cloned cDNA by ~1.5 kb. Because this first ATG triplet is not in the most favourable context for the initiation of translation¹⁷, and because only one clone

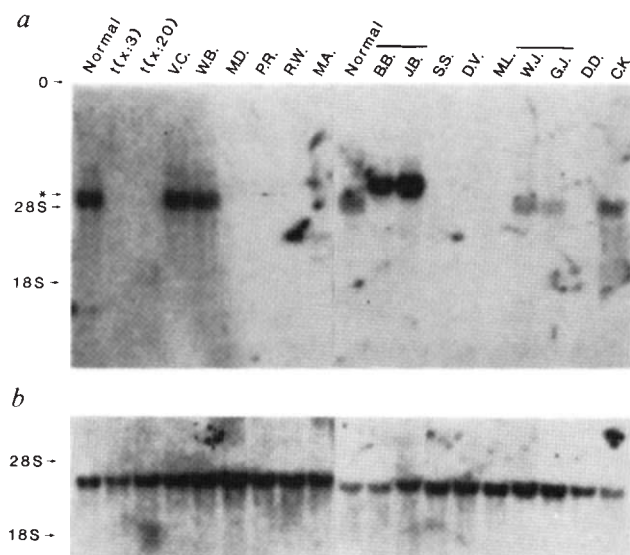


FIG. 2 The transcript detected by OCRL-1 is absent or abnormal in fibroblasts of OCRL patients. *a*, Northern blot analysis of fibroblast RNA using the insert of λ HK7 labelled by random priming as probe³¹. Arrows show the positions of the origin (O), the normal transcript (asterisk), and 28S and 18S ribosomal RNA markers. Symbols above each lane indicate control samples (Normal) two female translocations (X;3 and X;20) and the 15 male OCRL patients. A line across two lanes indicates that the two patients are brothers. *b*, The same filters as in *a* stripped of probe and rehybridized with a fragment of hSNF2L, a cDNA unrelated to OCRL which is located near to but entirely proximal to the X chromosomal translocation breakpoint in the t(X;3) female OCRL patient (I.O. *et al.*, manuscript submitted).

METHODS. Total RNAs were prepared from cultured fibroblasts by the guanidine isothiocyanate-caesium chloride method³¹. Northern blotting was done by standard procedures³¹ using 15 µg total RNA from each cell line separated on a 1.2% agarose-formaldehyde gel.

FIG. 3 Consensus nucleotide sequence of OCRL-1. Both strands of λ HK7 and λ HK0 were completely sequenced several times using Sequenase kits (United States Biochemical) with either dGTP or 7-deaza-dGTP according to manufacturer's protocols. The amino-acid sequence derived by manual translation is shown; stop codons are indicated by asterisks.

1	T CAGCCTGACATGGATGAAGTCTTTTGTGTTTAAAGACTTCTTAGTGATTGTTATAGG	1801	ACCAGCGACCACCAAGCGCTGTAGCGCCCTCTTCCATATTGGGGTGAAGGTGTGGATGAA
2	* ThrTrpMetLysPhePheValIlePheLysSerPheLeuSerValIleLysValIleLysValIle	598	ThrSerAspHisLysPheValIlePheLysSerPheLeuSerValIleLysValIleLysValIle
61	AGCCTCTGACAGACTCTCAGCTCCAGCTCCCGCTCCCGCTCCCGGCGCCGCGCC	1861	CGAAGGTACCGGAAAGCTTTGAAGATAGTGACGCATCATGGACAGATGGAAATGAC
121	SerLeuLeuAspLysSerGlnLeuProAlaIleProArgSerArgLeuProAlaProGlyAla	618	ArgArgTyrArgLysValIlePheGluAspSerValIleArgIleMetAspArgMetGluAsnSer
128	CGCGCGGAGCTGTCTCTCAACGACGACGCGAGGCTGGGTGGGTGGGGACGCGGG	1921	TTCTTCTCTCTTACAGACTCAGCAGGAGGCTTTGGTTTGAAGATGTGAAGTTTCGG
31	ArgArgGlyAlaValIleProGlnThrThrArgSerArgGlyIleTrpValIleTrpGlyArgGly	638	PheLeuProSerLeuGluLeuSerArgArgGlyPheValIlePheGluAsnValIleLysPheArg
181	AGCCAGTGTCTGCGGATCGGCGCGAGTCCGCTGCTGCTGAGCCGCGAGGCGCGCTGG	1981	CAACTACAAAGGGGAAGTTCAGCATCAGCAACAATGACAGGTTCCCTGCCATTTTCT
58	SerGlnCysArgArgIleGlyIleProGlnSerAlaValIleLeuLeuSerProGluAlaIlePheSer	658	GlnLeuGlnLysGlyLysPheGlnIleSerAsnAsnGlyIleValIleProCysHisPheSer
241	ATGGAGCCGCGCTCCCGGTGGAGCCGCGCTTGCACATGTCGAGGGTATGGAGATG	2041	TTCATCCCTAACTTAATGACAGCCAGTACTGCAAGCCATGGCTCGGGCTGAACCTTTT
78	MetGluProProLeuProValIleGlyAlaGlnProLeuAlaThrValIleGluIleMetGluMet	678	PheIleProLysLeuAsnAspSerGlnTyrCysLysProTrpLeuArgAlaGluProPhe
301	AAGGGTCTCTCCGGGAGCCCTGCGCCCTGACCTAGCCAGGAGCGCCGCAATATGAG	2101	GAGGCTACTTGGAGCCCAATGAGACAGTGGACATTTCTCTGTATGTATGTCAGCAA
98	LysGlyProLeuArgGluProCysAlaLeuThrLeuAlaGlnArgAsnGlyIleTyrGlu	698	GluGlyTyrLeuGluProAsnGluThrValIleSerLeuAspValIleTyrValIleSerLys
361	TTAATAATCCAGTTGATGAGAAGGACAGCATGTTCAAGATATCATCTTCAATAATAGC	2161	GACTCTGTAACTCTGAACCTGGGAGAGATAGATTGAAGATATCTCTGCTCTTAC
118	LeuIleIleGlnLeuHisGlnGlyIleGluGlnHisValIleGlnAspIleIleProIleLeuAsnSer	718	AspSerValIleThrIleLeuAsnSerGlyGluAspLysIleGluAspIleLeuValIleHis
421	CACCTTCAGATGTGTTCAGAAGCAGAAAGAACTCTTTGATTGACATAGCTTCAACAGT	2221	CTGGATCGAGGCAAGATTACTTCTTGACATCAGTGAAATTTACTCCCAAGTTGTTT
138	HisPheArgCysValIleGlnAlaGluGluThrLeuLeuIleAspIleAlaIleSerAsnSer	738	LeuAspArgGlyIleLysAspTyrPheLeuThrIleSerGlyAsnTyrLeuProSerCysPhe
481	GGCTGCAAAATTCGGGTTTCAGGGGACCTGGATTCAGAGACCAATTTCTCTGAATTTG	2281	GGCACATCTTAGAGGCTGTGCGGATGAGACCAATCTTCTGTATGTATGTCAGCAA
158	GlyCysLysIleArgValIleGlnGlyAspTrpIleArgGluArgArgPheGluIleProAsp	758	GlyThrSerLeuGluAlaLeuCysArgMetLysArgProIleArgGluValIleProValThr
541	GAGGAACACTGTTTGAAGTTCTCTCAGCTGCTCTGCTGCTCAGAAAGCTCAGTCACAG	2341	AAACTCAGACTGGAGAAATCCCTTCTGCAATGGTTCCTTTGGATGAAGGTGCCAT
178	GluGluHisCysLeuLysPheLeuSerAlaValIleLeuAlaGlnProLysAlaGlnSerGln	2398	LysLeuIleAspLeuGluLysSerAlaValIleLeuAlaGlnProLysAlaGlnSerGln
601	CTTCTGTTCAGAGCAAAAGGACTCATCTAGCTGTACCAAGAAATAGACACTAAGGAC	2401	GAGAGACCCCTCAGGTTCCTCAAGGAGATCTGGCTCTAGTAGATCACCTATTCAATATC
198	LeuLeuValIleProGluGlnLysAspSerSerSerTrpTrpGlnLysLeuAspThrLysAsp	798	GluArgProLeuGlnValIleProLysGluIleTrpLeuLeuValIleHisLeuPheLysTyr
661	AAACCTCTGTTTTCAGGCGCTCTGGATTCTGAAGACAAATTTCTCTGAATTTG	2461	GCCTGTCCAGCAGGAGGCTGTCCAGGACCTGGAAATGAGAGAGAGCTCCAGCAGATC
218	LysProSerValIlePheSerGlyLeuLeuGlyPheGluAspAsnPheSerSerMetAsnLeu	818	AlaCysHisGlnGlnGluAspLeuPheGlnThrProGlyMetGluGlnGluGlnGlnIle
721	GACAAGAAATAAATTCACAAATCAGCCTACTGGGATTCATGGGAACCCCACTCCA	2521	ATTGATTGTCTGGATACCAAGATTCCTGAGACAAATCCCTGGCAGCAACCTCTGTGGCT
238	AspLysLysIleAsnSerGlnAsnGlnProThrGlyIleHisArgGluProProProPro	838	IleAspCysLeuAspThrSerIleProGluThrIleProGlySerAsnHisSerValIle
781	CCCTTTTCAGTGAATAAAATCTCCAGCTGAAAAAGAGCTTCAACAGGAGCAGCC	2581	GAGGCACTGCTCATTTCTTGGAGGCTCCGACAGGAGCTCATCTGTACGAGCTGTAT
258	ProPheSerValIleAsnLysMetLeuGluGluGluGluGluGluGluGluGluGluGluGlu	858	GluAlaLeuLeuIlePheLeuGluAlaLeuProGluProValIleLysCysTyrGluLeuTyr
841	AAAGTGACCAACCACTTCAGGAGCTCTTTGACCAAAATACCAATCTGGGACGCGGAG	2641	CAGCGATGCTGTGACTCTGCTTATGATCCCCGATCTGCCAGAGGTGATCTCCAGCTT
278	LysValThrAsnThrMetArgLysLeuPheValIleProAsnThrGlnSerGlyIleGlnArgGlu	878	GlnArgCysLeuAspSerAlaIleTyrAspProArgIleCysArgGlnValIleIleSerGlnLeu
901	GGTCTCATCAACATATCTTGGCAAGCGAGAGAAAGATATGCAACATTCAGACTTTC	2701	CCGAGATGCCATAGAAATGTTTCGGTACTTGATGGCATTCCTTCGAGACTCTTAAAA
298	GlyLeuIleLysHisIleLeuAlaLysGluGluGluGluGluGluGluGluGluGluGluGlu	898	ProArgCysHisArgAsnValIlePheArgTyrLeuMetAlaIlePheLeuArgGluLeuLeuLys
961	AGATTGTTTGTGGAACTTGGAAATGTGAATGGCCAGCTCCAGATAGCGGGTGAAGACT	2761	TTCTTGAAATACAATAGCGTCAATGCCAATGATCGCTACTCTCTCTCACTAGCTTCTC
318	ArgPhePheValIleGlyThrTrpAsnValIleAsnGlyIleGlnSerProAspSerGlyLeuGluPro	918	PheSerGluTyrAsnSerValIleAsnAlaAsnMetIleAlaThrLeuPheThrLeuLeuLeu
1021	TGGCTGAACGTGTGATCCCAATCTCTGATATCTACTGCATGGATTCCAAGAACTGGAC	2821	CTGAGGCTCCACCAACCTTATGGCAAGACAGACTCCAAGTGACCCAGCGTGTCTATT
338	TrpLeuAsnCysAspProAsnProProAspIleTyrCysIleGlyPheGlnGluLeuAsp	938	LeuArgProProProAsnLeuMetAlaArgGlnThrProSerAspArgGlnArgAlaIle
1081	TGAGCAGCAAGCTCTCTCTTCTTGAATCTTGAAGCAACAAAGTTCCTGGCT	2881	CAGTCTCTTGGGCTTCTGCTTGGGAGCAAGAGACTTAAGCTTTTACTGTCTCTG
358	LeuSerThrGluAlaPhePheTyrPheGluSerValIleGlyGluGlnGluTrpSerMetAla	958	GlnPheLeuLeuGlyPheLeuLeuGlySerGluGluAsp *
1141	GTAGAGAGAGGTTTGATCTCCAAAGCAAGTATAAGAAAGTTCAACTGGTGGCCTTGT	2941	ATATTCTAGAAGCAGAGATCTCGGGCTCCAAGTATTTCAAGATGATTTAAAGTCATG
378	ValGluArgGlyLeuHisSerLysAlaLysTyrLysValIleGlnLeuValIleArgLeuVal	3001	CCACAGGAAGGGCTATTGCAAGATTTCAAGTTCTGTTTATAGTAAAGGAAGAGCGTT
1201	GGGATGATGCTTCTATATTGCCCAAGAGGATCAGTGTGATACATTCGTGATATTGCT	3061	TCTTAATCCCTCCTTACCATATCTCACAGAAAAATCACTTTAGACTTATATTGCCAA
398	GlyMetMetLeuLeuIlePheAlaArgLysAspGlnCysArgTyrIleArgAspIleAla	3121	GCCAAAGTTACCATATTTTGGTGTTTTGTGTTTCTCTTTATAAGGCAAAAGATCTGT
1261	ACAGAAACAGTTGGAACCTGGAATCATGGGAAAAATGGGAACAAAGGTGGGGTGAAGT	3181	ATTTACACTCCTTCACTAAGGATGTGTGTGTGCTCCCTACCAATTTGTATGATGT
418	ThrGluThrValIleGlyThrGlyIleMetGlyLysMetGlyAsnLysGlyIleValIleAlaVal	3241	CCTTAGTACCTTAGGCTAGATCTGAGATCTCCCATCTAGGCTCACAAGCACTACT
1321	AGATTGTTGTTTCAACAACCACTTTTGCATTTGCAATTCACATCTGGCTGCACAGTG	3301	CTGTAGCTGAGACTTGTCTAGAGTCTTGTGTTTGCATTTTGACCCACCTCTCTGCTG
438	ArgPheValIlePheHisAsnThrThrPheCysIleValIleAsnSerHisLeuAlaHisVal	3361	ATCACTCTTGTGACACTCCCGGAATTCCTGTCACTTTGAAGCAAGTCTGAGTGAGG
1381	GAGGACTTTGAGAGAAGGAATCAAGATTATAAGGACATTTGTGCGAGAATGAGTTTGTG	3421	CTAGTGACTCTTGGGTGCTTCAACAGTGAATTCAGTCTGCTGCTGACGTTATTACATG
458	GluAspPheGluArgArgAsnGlnAspTyrLysAspIleCysAlaArgMetSerPheVal	3481	CATTGTGCAATCTACTACAATGGCATCTTATGCTCTGTAACATTTGGCTTTTATGAG
1441	GTCCCAATTCAGACCTCCCGAGTGTGAACATCAGAACATGAGGTGTGCTTGGTGTG	3541	CTCCACATGGGTGGAAGGATATTCTTATAGATCAATTTAGTAGCATAACTGTAGGAC
478	ValProAsnGlnThrLeuProGlnLeuAsnIleMetLysHisGluValIleIleTrpLeu	3601	TATTAGAGATGGATCTCATGATGAGAGAGAATCAACATAGAAATGGAAGGATCTTGG
1501	GGAGATTGTAATATAGACTTTGATGCTGATGCCAATGAGGTGAAAGTCTTATTAAT	3661	TATCTGAAGAGTGAGAGACTTCATGTTTGAAGGCTTCTGCTTCCCATCTTCTCTG
498	GlyLysAspLeuGlnArgLysPheGluGlyGlyIleLysPheIleProThrTyrLysTyrAsp	3721	TTATTATCAAAATTTACAAAGGACTAATCTGGGTGCTCTGAGACCCCTCTGCTGCC
1561	AAGAAGACCTTCAGAGACTCTTGAATTCGACAGCTAAATATTCAGCGCACACAGAA	3781	TAGACATCCACCTCCAGAGCAACACTGGGCCACAGTAAAGAGGAGGTTGTACCTCA
518	LysLysAspLeuGlnArgLysPheGluGlyGlyIleLysAsnIleGlnArgThrGlnLys	3841	GGCAGGCCATCAGAGCTATGCTCTTCCACAGCAAGAGGATTTGGTGAAGCCCTT
1621	AAAGCTTTTGTGACTTCAATGAAGGGGAAATCAAGTTTCATCCCACTTATAGTATGAC	3901	GAATCTATCTCTGCTGCTTCTGAAATACCAAGGACAGATGCTCACTCTCTTCCAGAGG
538	LysAlaPheValIleAspPheAsnGluGlyGluIleLysPheIleProThrTyrLysTyrAsp	3961	ACTGACTCTGGGCTTCAACAGCTCTTCAACATAAGGGTTTAGAGACTCTCCCTTGGC
1681	TCTAAACAGACCGGTGGGATTCAGTGGGAAATGCCGGGTCCAGCCTGGTGTGACCGA	4021	TCCAGTCACCATATCCAGTGTGTGTGAAGAGCTGGCCACAGGACCAACCAAGCC
558	SerLysThrAspArgTrpAspSerSerGlyLysCysArgValIleProAlaTrpCysAspArg	4081	TTACCTTCCCACTACAGATGACCTTGAAGTTCATATTAATCAAGCTCTGTGTGACA
1741	ATCTCTTGGAGAGGAACAAATGTTAATCAGCTTAATATCGAGTCACATGGAATGAAA	4141	GCCTTTTATTAATAATTAATCATATATGTTGGTTGACAAACAGGACCAACCAAGTACT
578	IleLeuTrpArgGlyThrAsnValIleGlnLeuAsnTyrArgSerHisMetGluLeuLys	4201	GCAAACTGCTGTAGTGTGTTGGTCTCTGCTGTTTCTTTTGTGACAAACCTCTGCTG
		4321	CTGTTTACATTAATGCAAGGAGCAATACGCTGATGTACAACTCTGGGCACTG

extended far enough 5' to include the region of the in-frame stop at position 7, the authenticity of this ATG codon as the translation initiation codon is uncertain. Structural analysis, hydropathy calculations and searches of the ProSite database of protein motifs did not identify features of functional importance in OCRL-1 (ref. 18).

Searches of the current GenBank database with the BLAST (ref. 19) and FASTA (ref. 20) algorithms revealed 53% amino-acid identity and 71% similarity (Fig. 4) between OCRL-1 and human inositol polyphosphate-5-phosphatase (accession no. M74161), a partial cDNA sequence for a soluble 75K inositol polyphosphate-5-phosphatase from human platelets²¹. The similarity extends throughout the entire 672-amino-acid sequence of HUMINP5P and is not localized to a particular domain. The significant differences in predicted amino-acid sequence and relative molecular mass, however, indicate that OCRL-1 is not HUMINP5P, but it could represent the gene for a 100K inositol polyphosphate-5-phosphatase purified from brain^{22,23}.

The discovery of this strong homology between HUMINP5P

and OCRL-1 is relevant to the normal physiology and pathology of the isomers and metabolites of inositol phosphate and to the pathological effects of abnormalities in their metabolism. Inositol(1,4,5)-trisphosphate is involved in signal transduction and phosphatidyl inositol participates in anchoring membrane proteins by glypiation; however, the function of many inositol phosphates, either free or esterified in phospholipids, is unclear^{24,25}. On the basis of the strong amino-acid sequence homology between OCRL-1 and HUMINP5P, we propose that OCRL is an inborn error of metabolism of one or more inositol phosphate metabolites. If so, OCRL would be the first human genetic disorder affecting these pathways and would constitute a genetic system in which to explore the normal physiological functions of these intermediates and the effect of abnormalities in their pathways on human development. Abnormalities in inositol metabolism or transport have been implicated in the pathogenesis of the cataracts and peripheral neuropathy in galactosaemia and diabetes mellitus²⁶⁻²⁸. Identification of the gene responsible for OCRL may help elucidate the biochemical

OCRL-1 247 PTGIHREPPPPFVSNKMLPREKEASNEQPKVTNTMRKLFVPTQSGOR
 HUMINP5P 1 PESLQPRQNKSKSEITDMVRSSTITVSDKAHILSMQKFLGR
 OCRL-1 297 EGLIKHILAKREKEYVNIQTFRFFVGTNVNNGQSPDSGLPEWLNCDPNP
 HUMINP5P 42 DTIVKSHLLKQEDDYTIQNFRRFAGTYNVNNGQSPKCECLRLWLSNGIQAP
 OCRL-1 347 DIYICIGFQELDLSTEAFFYFESVKEQWMAVERGLHSAKAKYKQLVRL
 HUMINP5P 92 DVYCVGFQELDLKSAFFHDTPEKEEFKAVSEGLHPDAKYAKVLIRL
 OCRL-1 397 VGMMLLIFARKDQCRYIRDIATETVGTGIMGMKMGKGVAVRFVHNTTF
 HUMINP5P 142 VGMMLLLVYKGAHAAYISEVEAETVGTGIMGRMGKMGVAVRFQFHNTSI
 OCRL-1 447 CIVNSHLAAHVEDFERRNQDYKDCARMSFVVPNTQLPQLNIMKEVVIW
 HUMINP5P 192 CVVNSHLAAHIEYERRNQDYKDCSRMQFCQDPPLPLTISNDHVLW
 OCRL-1 497 LGDLNRYLCMPDANEVKSILKKDLQRLKFDQLNIQRTQKAFVDFNEG
 HUMINP5P 242 LGDLNRYLEELDVEKVKLLIEKDFQMLYAYDQLKIQVAAKTVFEGFTEG
 OCRL-1 547 EIKFIPTYKYDSKTRDWSGSKCRV ..PAWCDRLVRGTNNVNLNRYSHM
 HUMINP5P 292 ELTFQPTKYDTRALTGTIPVRSVALLPGVIG.FLWKGKNITQLSYQSHM
 OCRL-1 597 ELKTSOHKPVSAFHFVIGVVDERRRYKVFDSVRIMDRMENDPLSLEL
 HUMINP5P 341 ALKTSOHKPVSVFDIGVRVNDLRYKTLLEEIVRSLOKMEANIPSVSL
 OCRL-1 647 SRREFVFNKFRQLQKGFQISNNGQVPCFHSFIPKLNDSQYCKPWLRA
 HUMINP5P 391 SKREFCFQNVKYQMLKVESFTI..HNGQVPCFHFINKPDEESYCKQWLNA
 OCRL-1 697 EPFEGYLEPNETVDISLDVYVSKDSVTILNSGEDIEDILVHLDRGKDY
 HUMINP5P 440 NPSRGFLLPDSVDEIDLEL FVNKTATKILNSGEDIEDILVHLDRGKDY
 OCRL-1 747 FLTISGNYLPSCFGSLALCRMRKPIREVPVTKLIDLEKSLQMVPLDE
 HUMINP5P 490 FLSVSGNYLPSCFGSPHITLCYMRPILOLPLETISEL..TLMPPVWTGDD
 OCRL-1 797 GAS.ERLPQVPKEIWLVDHLYKAYHQEDLFQTPGMQELQQIIDCLDT
 HUMINP5P 538 GSQLDSPMEIPKELVMMVDYLYRVAQVEDLFQPGGLRSEFHIRDCLDT
 OCRL-1 846 SIPETIPGSNHSVAEALLIFLEALPEPVICYELYQRCLOSAIDPRICQV
 HUMINP5P 588 GMIDLSASNHSVAEALLIFLESLEPVICYSTYHNCLECSPHYTASKQV
 OCRL-1 896 ISQLPRCHRVFRYLMALFRELLEKFSEYNSVYNNM
 HUMINP5P 638 ISTLPFHKNVFHYLMALFRELLEKNSAKNHLNDEI

FIG. 4 Alignment of predicted amino-acid sequence of OCRL-1 with the predicted sequence of human platelet inositol polyphosphate-5-phosphatase (HUMINP5P). Amino-acid residues are shown in single-letter code. Vertical lines indicate identity, double dots indicate highly homologous substitutions (similarity score between 0.5 and 1.5), and single dots represent less homologous substitutions (similarity score between 0.2 and 0.5). Similarity was established according to a normalized Dayhoff matrix³².

basis for these more common disorders and contribute to developing new strategies for their management. □

Received 6 April; accepted 19 May 1992.

- Lowe, C. V., Terrey, M. & MacLachan, E. A. *Am. J. Dis. Child.* **83**, 164-184 (1952).
- McKusick, V. A. in *Mendelian Inheritance in Man* (8th edn) 1335-1336 (Johns Hopkins Univ. Press, Baltimore, 1989).
- Charnas, L. R., Bernardini, I., Rader, D., Hoeg, J. & Gahl, W. A. *New Engl. J. Med.* **324**, 1318-1325 (1991).
- Yamashina, I., Yoshida, H., Fukui, S. & Funakoshi, I. *Molec. cell. Biochem.* **52**, 107-124 (1983).
- Kieras, F. J., Houck, G. E., French, J. H. & Wisniewski, K. *Biochem. Med.* **31**, 201-210 (1984).
- Harper, G. S., Hascall, V. C., Yanagashita, M. & Gahl, W. A. *J. Biol. Chem.* **262**, 5637-5643 (1987).
- Yoshida, H., Fukui, S. & Yamashina, I. *Biochem. biophys. Res. Commun.* **107**, 1144-1150 (1982).
- Horie, K., Yano, T., Funakoshi, I. & Yamashina, I. *Clin. chim. Acta* **177**, 41-48 (1988).
- Silver, D. N., Lewis, R. A. & Nussbaum, R. L. *J. Clin. Invest.* **79**, 282-285 (1987).
- Reilly, D. S., Lewis, R. A., Ledbetter, D. H. & Nussbaum, R. L. *Am. J. Hum. Genet.* **42**, 748-755 (1988).
- Reilly, D. S., Lewis, R. L. & Nussbaum, R. L. *Genomics* **8**, 62-70 (1990).
- Hodgson, S. V. *et al. Am. J. Med. Genet.* **23**, 837-847 (1986).
- Mueller, O. T. *et al. Am. J. Hum. Genet.* **49**, 804-810 (1991).
- Nelson, D. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 6157-6161 (1991).
- Okabe, I., Attree, O., Bailey, L. C., Nelson, D. L. & Nussbaum, R. L. *J. Inher. Metab. Dis.* (in the press).
- Wallace, M. R. *et al. Science* **249**, 181-186 (1990).
- Kozak, M. *Nucleic Acids Res.* **15**, 8125-8132 (1987).
- Devereux, J., Haeblerli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387-395 (1984).
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. *J. molec. Biol.* **215**, 403-410 (1990).
- Pearson, W. R. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2444-2448 (1988).
- Ross, T. S., Jefferson, A. B., Mitchell, C. A. & Majerus, P. W. *J. Biol. Chem.* **266**, 20283-20289 (1991).
- Hansen, C. A., Johanson, R. A., Williamson, M. T. & Williamson, J. R. *J. Biol. Chem.* **262**, 17319-17326 (1987).

- Erneux, C. *et al. Eur. J. Biochem.* **181**, 317-322 (1989).
- Rana, R. S. & Hokin, L. E. *Physiol. Rev.* **70**, 115-164 (1990).
- Majerus, P. W. *et al. Cell* **63**, 459-465 (1990).
- Cammarata, P. R., Tse, D. & Yorio, T. *Curr. Eye Res.* **9**, 561-568 (1990).
- Cammarata, P. R., Tse, D. & Yorio, T. *Diabetes* **40**, 731-737 (1991).
- Clements, R. S. *Adv. exp. med. Biol.* **124**, 105-131 (1979).
- Scott, A. F., Phillips, J. A. & Migeon, B. R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4563-4565 (1979).
- Nussbaum, R. L., Airhart, S. D. & Ledbetter, D. H. *Am. J. Med. Genet.* **23**, 457-466 (1986).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1989).
- Gribskov, M. & Burgess, R. R. *Nucleic Acids Res.* **14**, 6745-6763 (1986).

ACKNOWLEDGEMENTS. We thank V. Dubowitz, T. Mueller, M. Palmieri, R. Hofnagel, M. Ampola, J. Schneider and D. Whelan for cell lines from Lowe syndrome patients; L. Johnson, N. Blanchard, D. Sosnoski and M. F. Victoria for technical assistance; and R. Taub, G. Dreyfuss, S. Liehaber and T. Kadesch for critically reading the manuscript. This work was supported in part by grants from NICH, by a French Armed Forces' Scholarship to O.A., a fellowship from the Life and Health Insurance Medical Research Fund to L.C.B., grants to R.A.L. from the Lowe Syndrome Association (West Lafayette, Indiana) and Research to Prevent Blindness (New York), and by grants from the Department of Energy and the Hereditary Disease Foundation to D.L.N.

Nuclear localization and signalling activity of phosphoinositidase C_β in Swiss 3T3 cells

Alberto M. Martelli*, R. Stewart Gilmour†, Valeria Bertagnolo‡, Luca M. Neri‡§, Lucia Manzoli|| & Lucio Cocco*¶

Institutes of Human Anatomy at the Universities of * Bologna, V. Irnerio, 48, 40126, ‡ Ferrara, V. Fossato di Mortara, 66, 44100 and || Chieti, V. dei Vestini, 66013, Bologna, Italy
 § Institute of Cytomorphology, Consiglio Nazionale delle Ricerche, c/o Istituto Rizzoli, Bologna 40136, Italy
 † Department of Molecular and Cellular Physiology, AFRC, Institute of Animal Physiology and Genetics Research, Babraham, Cambridge CB2 4AT, UK

THE hydrolysis of phosphatidylinositol 4,5-bisphosphate (PtdInsP₂) is a widespread receptor-coupled signalling system at the plasma membrane of most eukaryotic cells. The existence of an entirely separate nuclear phosphoinositide signalling system is suggested from evidence that purified nuclei synthesize PtdInsP₂ and phosphatidylinositol 4-phosphate (PtdInsP) *in vitro*¹ and that a transient decrease in the mass of these lipids occurs when Swiss 3T3 cells are cultured in the presence of insulin-like growth factor-1 (IGF-1)²⁻⁴. These IGF-1-dependent changes in inositol lipids coincide with an increase in nuclear diacylglycerol⁴ and precede translocation to the nucleus and activation of protein kinase C (refs 5, 6). Circumstantial evidence that links these changes with mitosis comes from the isolation of a 3T3 clone that expresses the type-1 IGF receptor and binds IGF-1 peptide but does not respond mitogenically or show transient mass changes in nuclear inositol lipids⁷. A key question is how IGF-1 initiates the rapid breakdown of PtdInsP and PtdInsP₂ in the nucleus. Here we present evidence that nuclei of 3T3 cells contain the β-isozyme of phosphoinositidase C, whereas the γ-isozyme is confined to the cytoplasm and that IGF-1 treatment stimulates exclusively the activity of nuclear phosphoinositidase C.

An analysis of phosphoinositidase C (PIC) activities in the nucleus and cytoplasm of Swiss 3T3 cells using [³H]PtdInsP and [³H]PtdInsP₂ as substrates is shown in Fig. 1. Both nuclear and cytoplasmic activities require calcium because the addition of EGTA completely abolishes breakdown. This argues against the presence in Swiss 3T3 cells of the α form of PIC which is insensitive to calcium with PtdInsP and PtdInsP₂ as substrates⁸. Near-maximal responses are obtained in the absence of exogenous calcium, suggesting that the nanomolar levels of calcium present in the nuclear and cytoplasmic compartments⁹ are sufficient for activity. The presence of exogenous calcium

¶ To whom correspondence should be addressed.

FIG. 1 Determination of PIC enzymic activity and properties in isolated nuclei and cytoplasmic fractions from Swiss 3T3 cells. *a, b*, Calcium and pH dependency, respectively. Results are the average of five separate experiments ($\leq 8\%$ for each point). *c, d*, Representative HPLC profiles of inositol phosphates liberated by nuclear PIC using [^3H] PtdInsP (*c*) and [^3H] PtdInsP₂ (*d*) as substrates. Inositol 1,4-phosphate (InsP₂) from cytoplasmic fraction (■); Inositol 1,4,5-phosphate (InsP₃) from cytoplasmic fraction (●); InsP₂ from nuclei (*); InsP₃ from nuclei (+).

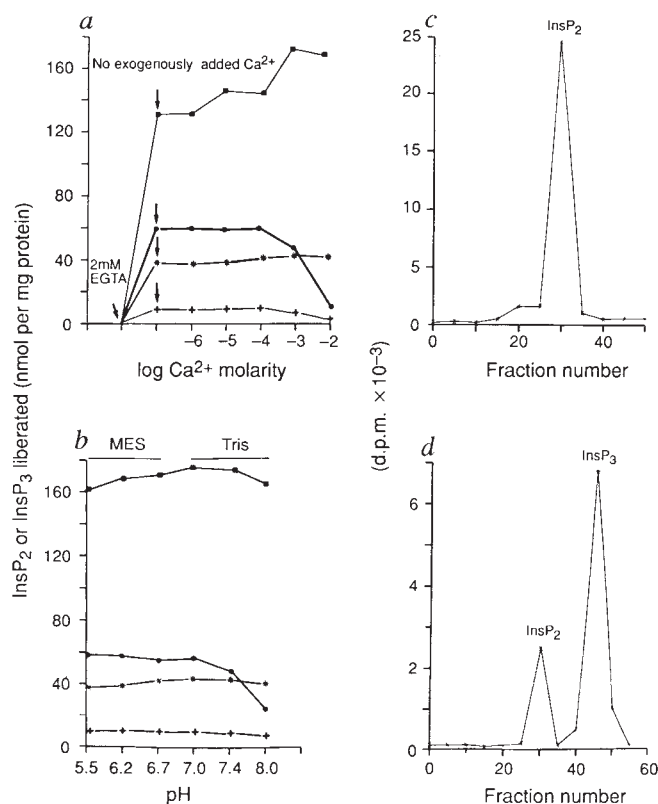
METHODS. To obtain nuclei, cells (5×10^6) were suspended in 500 μl of 10 mM Tris-HCl, pH 7.8, 1% Nonidet P-40, 10 mM β -mercaptoethanol, 0.5 mM phenylmethylsulphonyl fluoride, 1 $\mu\text{g ml}^{-1}$ aprotinin and leupeptin and 10 $\mu\text{g ml}^{-1}$ soybean trypsin inhibitor for 2 min at 0 °C. Then 500 μl double-distilled H₂O was added and the cells allowed to swell for 2 min. Cells were sheared by 10 passages through a 22-gauge needle. Nuclei were recovered by centrifugation at 400g for 6 min and washed once in 10 mM Tris-HCl, pH 7.4, 2 mM MgCl₂, plus protease inhibitors. In some cases calpain inhibitor-1 (15 $\mu\text{g ml}^{-1}$) and -2 (7 $\mu\text{g ml}^{-1}$) were used (Boehringer). The cytoplasmic fraction was obtained by homogenizing cells with 20 strokes in a Dounce homogenizer in 10 mM Tris-HCl, pH 7.8, 10 mM β -mercaptoethanol and protease inhibitors and then pelleting the nuclei at 400g. PIC assays (100 μl) contained 100 mM MES or Tris buffer, 150 mM NaCl, 0.06% taurodeoxycholate, 3 nmol [^3H] PtdInsP or [^3H] PtdInsP₂ (specific activity 30,000 d.p.m. nmol^{-1} ; Amersham), 60 μg nuclear or cytoplasmic protein. Incubation was at 37 °C for 30 min. The time course of reaction (not shown) was linear for 30 min before reaching a plateau. Hydrolysis was stopped by adding chloroform-methanol-HCl as described in ref. 4, and inositol phosphates recovered from the aqueous phase were analysed by HPLC essentially as described in ref. 27. Samples were loaded onto a Partisil 10 SAX column and eluted with a linear gradient from distilled water to 2 M ammonium formate (pH 3.7, adjusted with phosphoric acid) at 1.5 ml min^{-1} . As reference compounds, [^3H] InsP and [^3H] InsP₂ (100,000 d.p.m.) were used and eluted as described. Fractions (1 ml) were collected, dissolved in Hionic-fluor (Packard) and counted by liquid scintillation. Protein was measured as in ref. 28. When PtdInsP₂ is used as substrate, the HPLC elution profile also shows the presence of some InsP₂: because the ^3H labelling is in the inositol ring, InsP₂ may be generated by a monoesterase activity on the inositol moiety of the substrate.

up to 100 μM had little further effect with either substrate; however, higher calcium levels caused a progressive inhibition of both nuclear and cytoplasmic PtdInsP₂ phosphoinositidase.

Both nuclear and cytoplasmic enzymes act optimally over a broad pH range, however the latter shows greater instability above pH 7.0. The ratio of maximal activities for the two substrates also differs for the two fractions. The ratio of PtdInsP₂:PtdInsP breakdown is 1:2.8 and 1:4, respectively, for the cytoplasmic and nuclear enzymes.

FIG. 2 PIC activity in isolated Swiss 3T3 nuclei and cytoplasmic fractions after stimulation with IGF-1. *a*, [^3H] PtdInsP as substrate; *b*, [^3H] PtdInsP₂ as substrate. Black bars, cytoplasmic fraction; white bars, isolated nuclei. *c*, Inhibition of solubilized nuclear PIC activity by anti-PIC- β , - γ and - δ antibodies; *d*, inhibition of cytoplasmic PIC activity by anti-PIC- β , - γ and - δ antibodies. Black bars, anti- γ ; white bars, anti- β ; hatched bars, anti- δ antibodies. Results are the average of five separate experiments (s.d. $\leq 8\%$ for each point).

METHODS. Swiss 3T3 cells were grown exactly as described in ref. 2. Briefly, cells grown for 6 days in Dulbecco's modified Eagle's medium (DMEM) containing 10% newborn bovine serum were washed twice with serum-free DMEM containing 1% bovine serum albumin and then incubated in the same serum-free medium containing 20 ng ml^{-1} IGF-1 for the time indicated. Assays were carried out at the optimum pH and calcium concentrations (see Fig. 1). Solubilization of nuclear PIC activity. Nuclei prepared as described in Fig. 1 were suspended at 1 mg DNA ml^{-1} in 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA and allowed to swell at 4 °C for 10 min before rupturing by 40 passages through a 25-gauge hypodermic needle. The lysate was centrifuged for 10 min at 48,000g and the PIC activity of supernatant determined by the standard assay. In these conditions there was a threefold increase in PIC-specific activity between nuclei from control and 2-min-stimulated cells. For inhibition of PIC activity by isozyme-specific antibodies, 5 μg nuclear lysate protein was incubated at room temperature in PBS containing 0.65% BSA and 12.5 μg of either anti-PIC- β , - γ or - δ antibodies as described previously¹⁰. The cytoplasmic fraction was prepared as for Fig. 1 and 5 μg protein assayed under the same conditions.



Within 2 min of treatment of 3T3 cells with IGF-1 there is a transient and roughly 100% increase in nuclear PIC activity. After 30 min, nuclear PIC activity returns to control values (Fig. 2a, b). Because there are multiple forms of PIC (refs. 10-12), the assays were repeated after earlier treatment of nuclear and cytoplasmic fractions with neutralizing monoclonal antibodies against the β , γ and δ PIC isozymes¹⁰. For these experiments, nuclear PIC was solubilized by hypotonic shock (Fig. 2 legend), resulting in a quantitative release of all the original activity.

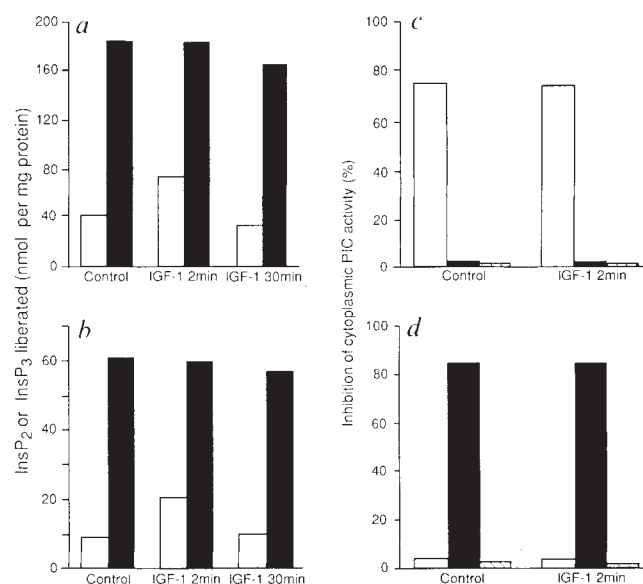
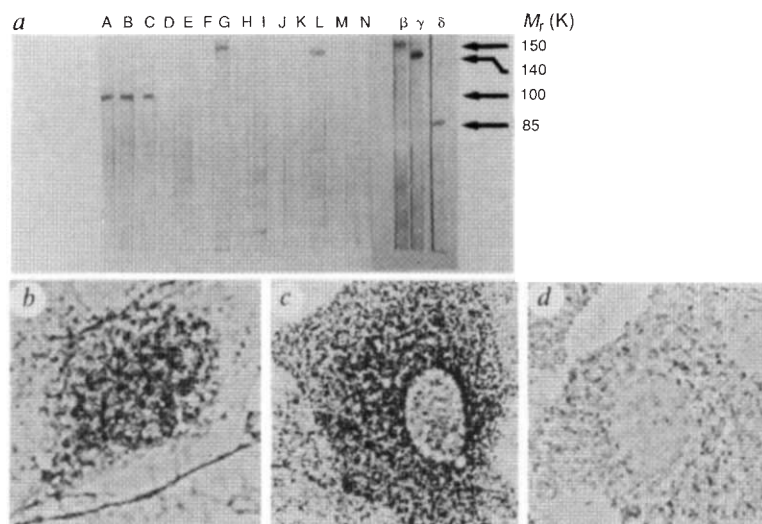


FIG. 3 Immunochemical and immunocytochemical detection of PIC isoforms in Swiss 3T3 fibroblasts. *a*, Western blot. Lanes β , γ and δ show the detection of the β -, γ - and δ -PIC isoforms (150, 140 and 85K, respectively) in a whole-cell extract of PC12 cells. In separate experiments it has been found that PIC- β locates almost entirely in the nucleus of these cells (S. Capitani *et al.*, manuscript in preparation). Lanes A–J, purified nuclei; K–N, cytoplasmic fractions from Swiss 3T3 cells. A, D–J, control nuclei; B, nuclei from 2 min IGF-I-stimulated cells; C, nuclei from 30 min IGF-I stimulated cells. A, B, C, G, K, Anti- β -isoform antibody; D, H, L, anti- γ -isoform antibody; E, I, M, anti- δ -isoform antibody; F, J, N, secondary antibody only. In samples A–F, nuclei were isolated in the absence of calpain inhibitors (Fig. 1 legend), whereas in samples G–J, inhibitors were present. *b*, *c*, *d*, Immunocytochemical analyses of PIC isoforms β , γ and δ , respectively, in Swiss 3T3 cells. Note that the antibody to PIC β reacts only with the nucleus and that the antibody to PIC γ reacts exclusively with the cytoplasm. Neither nucleus nor cytoplasm produced a reaction to the antibody to PIC δ (magnification, $\times 100$).

METHODS. Proteins (60 μ g) from purified nuclei and cytoplasmic fractions were separated on 8% polyacrylamide 0.1% SDS gels and transferred to nitrocellulose paper as in ref. 5. Strips were incubated with 1 μ g ml⁻¹ antibody to PIC isoforms (UBI Lake Placid, New York) according to the protocols described in ref. 10. Antibodies were



detected by alkaline phosphatase-conjugated anti-mouse IgG as described in ref. 5. Antibodies were used at the same concentrations as for the western blots. Staining procedures were as in ref. 5.

Using the same assay, 75% of the soluble nuclear PIC activity was inhibited by anti- β antibody, whereas the effects of anti- γ and anti- δ antibodies were negligible (<3%) (Fig. 2*c, d*). Cytoplasmic PIC, by contrast, was inhibited by 85% with anti- γ antibody and less than 4% with anti- β and - δ antibodies. The specificity of the antibodies was tested by western blotting using a whole-cell extract of PC12 cells which contain all three (β , γ and δ) PIC isozymes¹³. Figure 3 (right-hand lanes) shows the expected bands at 150K for PIC- β , 140K for PIC- γ and 85K for PIC- δ . Further confirmation of the distribution of individual isozymes in Swiss 3T3 cells was determined by western blotting (Fig. 3, lanes A–N).

A 100K nuclear protein was detected with anti- β antibody, which shifted to 150K when nuclei were purified in the presence of calpain inhibitors. This indicates that PIC- β resides in the nucleus and that the 100K enzyme is as catalytically active as the 150K enzyme. The conversion of the 150K to the 100K species of this isoform by the proteolytic action of calpain has been reported previously¹¹. Using antibodies to the other forms it was found that PIC- δ is absent in both nuclear and cytoplasmic fractions and that PIC- γ is only found in the cytoplasm. Separate confirmation was obtained from immunocytochemical analysis (Fig. 3*b*), which localized the PIC- β to the nuclear region, whereas the γ isozyme (Fig. 3*c*), as previously found in rat embryo fibroblasts¹⁴, is distributed throughout the cytoplasm. The immunoblot analysis also shows that the PIC- β is already present in the nucleus and suggests that the increased hydrolysis of nuclear PtdInsP and PtdInsP₂ induced by IGF-1 represents activation of pre-existing enzyme. The possibility that IGF-1 causes a translocation of cytoplasmic PIC- β to the nucleus seems less likely for the following reasons: IGF-1 treatment has no effect on the total PIC activity of the cytoplasm (Fig. 2*a, b*); western blot and immunocytochemical analysis with anti- β antibody are negative for cytoplasm (Fig. 3*a*, lane K, and Fig. 3*b*) and shows no changes in nuclear immunoreactivity after stimulation with IGF-1 (Fig. 3*a*, lanes A and B); if we interpret the small (4%) inhibition of cytoplasmic PIC activity by anti- β antibody (Fig. 2*d*) as representing a minor cytoplasmic presence of the β -isozyme rather than experimental background, it can be calculated that even if all of it were translocated to the nucleus it would be insufficient to account for the magnitude of the nuclear increase.

To eliminate possible artefacts due to cytoplasmic contamination all nuclear preparations were examined by electron

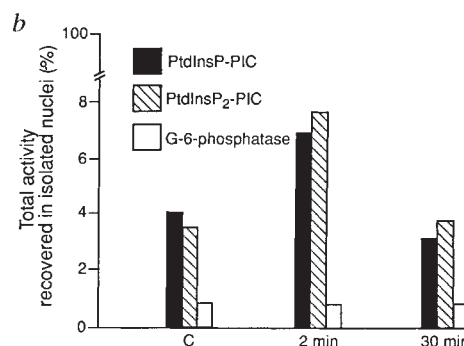
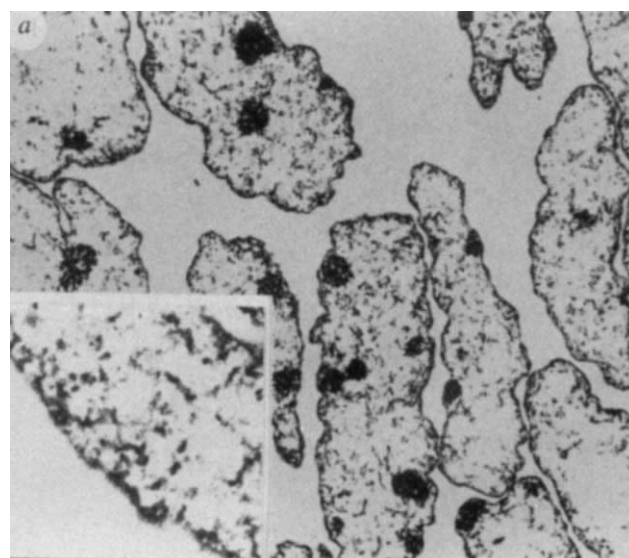


FIG. 4 *a*, Electron micrograph of nuclei isolated from Swiss 3T3 cells. Inset shows a higher magnification of the complete stripping of the nuclear envelope. Scale bar, 1 μ m. *b*, Percentage recovery of PIC and G-6-phosphatase activities in isolated nuclei.

METHODS. Samples from nuclear preparations (as described in Fig. 1) were processed for electron microscopy as in ref. 1. G-6-phosphatase was assayed as described in ref. 15. Activities were calculated by comparing the total enzymic activity of a given amount of cells to an equivalent amount of nuclei, taking into account both cell number and nuclear DNA content.

microscopy and shown to be completely stripped of their nuclear membranes and free from cytoplasmic debris (Fig. 4a). An additional proof of purity is that in our nuclear preparations the activity of glucose-6-phosphatase, a cytoplasmic marker¹⁵, also located at the outer nuclear envelope¹⁶, represents less than 0.8% of the total cellular activity, whereas PIC activity in control nuclei is ~4% (Fig. 4). Moreover, during the time course of treatment with IGF-1, G-6-phosphatase activity remains constant, whereas the nuclear PIC activity roughly doubles after 2 min of stimulation and then returns to control levels after 30 min.

PIC- β is regulated by the G_Q class of G proteins and in membrane preparations it can be stimulated by specific ligands that act through InsP₃ (refs. 17–21). The finding that in Swiss 3T3 cells PIC- β is confined principally to the nucleus is therefore

surprising. There are a variety of PIC- β isoforms found in different cells²² and it is conceivable that in 3T3 cells there are distinct cytoplasmic and nuclear enzymes and that only the latter is recognized by the anti- β antibody used here. Alternative and less likely explanations would require that the 3T3 cell was unique in having no involvement of PIC- β in cytosolic calcium regulation. Although this issue remains unresolved, we feel that in 3T3 cells there is good evidence for a separate nuclear PIC- β isoform whose activity is IGF-1-sensitive. It is not known with which nuclear structures PIC- β associates, but G proteins have been reported to be integral components of the nuclear lamina^{23,24}. These data strengthen recent arguments for the role of a discrete nuclear polyphosphoinositide signalling system which is quite distinct from that of the plasma membrane^{25,26}. □

Received 10 February; accepted 26 May 1992.

- Cocco, L., Gilmour, R. S., Ognibene, A., Manzoli, F. A. & Irvine, R. F. *Biochem. J.* **248**, 765–770 (1987).
- Manzoli, F. A. *et al. Adv. Enz. Reg.* **28**, 25–34 (1989).
- Cocco, L. *et al. Biochem. biophys. Res. Commun.* **159**, 720–725 (1989).
- Divecha, N., Banfic, H. & Irvine, R. F. *EMBO J.* **10**, 3207–3214 (1991).
- Martelli, A. M. *et al. Biochem. biophys. Res. Commun.* **177**, 480–487 (1991).
- Martelli, A. M., Gilmour, R. S., Falcieri, E., Manzoli, F. A. & Cocco, L. *Expl. Cell Res.* **185**, 191–202 (1989).
- Martelli, A. M. *et al. FEBS Lett.* **283**, 243–246 (1991).
- Bennet, C. F. & Crooke, S. T. *J. Biol. Chem.* **262**, 13789–13797 (1987).
- Tucker, R. W. & Fay, F. S. *Eur. J. Cell Biol.* **51**, 120–127 (1990).
- Suh, P. G., Ryu, S. H., Choi, W. C., Lee, K. Y. & Rhee, S. G. *J. Biol. Chem.* **263**, 14497–14504 (1988).
- Rhee, S. G., Suh, P. G., Ryu, S. H. & Lee, S. Y. *Science* **244**, 546–550 (1989).
- Bansal, V. S. & Majerus, P. W. *A. Rev. Cell Biol.* **6**, 41–67 (1990).
- Kim, U-H. *et al. J. Biol. Chem.* **266**, 1359–1362 (1991).
- McBride, K., Rhee, S. G. & Jaken, S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7111–7115 (1991).
- Garland, R. C. & Cori, C. F. *Biochemistry* **11**, 4712–4718 (1972).

- Thorne-Tjomsland, G., Clermont, Y. & Tang, X. *Biol. Cell* **71**, 33–41 (1991).
- Smrcka, A. V., Hepler, J. R., Brown, K. O. & Sternweis, P. C. *Science* **251**, 804–807 (1991).
- Taylor, S., Chae, H., Rhee, S. & Exton, J. *Nature* **350**, 516–518 (1991).
- Wu, D., Lee, C. H., Rhee, S. G. & Simon, M. I. *J. Biol. Chem.* **267**, 1811–1817 (1992).
- Shenker, A., Goldsmith, P., Unson, C. & Spiegel, A. *J. Biol. Chem.* **266**, 9309–9313 (1991).
- Gutowksi, S. *et al. J. Biol. Chem.* **266**, 20519–20524 (1991).
- Ryu, S. H. *et al. J. Biol. Chem.* **265**, 17941–17945 (1990).
- Rubins, J. B., Benditt, J. O., Dickey, B. F. & Riedel, N. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7080–7084 (1990).
- Seydel, U. & Gerace, L. *J. Biol. Chem.* **266**, 7602–7608 (1991).
- Michell, R. H. *Curr. Biol.* **2**, 200–202 (1992).
- Payastre, B. *et al. J. Biol. Chem.* **267**, 5078–5084 (1992).
- Heslop, J. P., Blakely, D. M., Brown, K. D., Irvine, R. F. & Berridge, M. J. *Cell* **47**, 703–709 (1986).
- Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. *J. Biol. Chem.* **193**, 265–275 (1951).

ACKNOWLEDGEMENTS. We thank Professor F. A. Manzoli for support and encouragement, R. F. Irvine for purified inositol lipids and inositol phosphate standards and A. M. Billi for assistance. This work was supported by grants from the Italian CNR PFG/PFACRO.

TCP1 complex is a molecular chaperone in tubulin biogenesis

Michael B. Yaffe[†], George W. Farr^{*}, David Miklos[‡],
Arthur L. Horwich[‡], Mona L. Sternlicht^{*}
& Himan Sternlicht^{*§}

^{*}Department of Pharmacology, Case Western Reserve School of Medicine, Cleveland, Ohio 44106, USA

[†]Department of Surgery, Harvard Medical School, New England Deaconess Hospital, Boston, Massachusetts 02146, USA

[‡]Howard Hughes Medical Institute and Department of Genetics, Yale School of Medicine, New Haven, Connecticut 06510, USA

A ROLE in folding of newly translated proteins in the cytosol of eukaryotes has been proposed for t-complex polypeptide-1 (TCP1), although its molecular targets have not yet been identified^{1–3}. Tubulin is a major cytosolic protein whose assembly into microtubules is critical to many cellular processes^{4–8}. Although numerous studies have focused on the expression of tubulin^{9–20}, little is known about the processes whereby newly translated tubulin subunits acquire conformations that enable them to form α - β -heterodimers. We examined the biogenesis of α - and β -tubulin in rabbit reticulocyte lysate, and report here that newly translated tubulin subunits entered a 900K complex in a protease-sensitive conformation. Addition of Mg-ATP, but not nonhydrolysable analogues, released the tubulin subunits as assembly-competent protein with a conformation that was relatively protease-resistant. The 900K complex purified from reticulocyte lysate contained as its major constituent a 58K protein that cross-reacted with a monoclonal antiserum against mouse TCP1. We conclude that TCP1 functions as a cytosolic chaperone in the biogenesis of tubulin.

β -Tubulin polypeptides synthesized in rabbit reticulocyte lysate are found in three molecular forms separable by anion-

exchange chromatography (Mono-Q): an early eluting form (β I) corresponding to a complex between β -tubulin and a reticulocyte protein(s), and forms eluting later that correspond to a free β -subunit (β II) and tubulin $\alpha\beta$ heterodimer (β III) (Fig. 1a)^{21,22}. These forms were further examined in pulse radiolabelling experiments (Fig. 1b). After synthesis for 10 min, β -tubulin was found almost exclusively in the β I form whereas β II was the major product after synthesis for 60 min (Fig. 1a)²¹. Pulse-chase experiments confirmed that β I was a precursor of β II and β III (Fig. 1c, d). To test whether the β -subunits released from β I were competent for assembly, pulse-chase studies were done in the absence and presence of bovine microtubule protein (MTP) (Fig. 2a). Because the endogenous tubulin concentration in reticulocyte lysate is low²³ in the absence of MTP, only a small amount of the β III heterodimer could be observed (Fig. 2a, '–MTP'). When exogenous MTP was supplied to increase the free α -tubulin pool, the major form observed was β III, indicating that the released β -subunits were indeed capable of associating with α -subunit (Fig. 2a, +MTP). These β III heterodimers were fully competent to assemble into microtubules in a stringent coassembly assay with MTP^{21,24} (data not shown).

Using limited proteolysis with chymotrypsin (Fig. 2b)^{21,25}, we investigated whether the conformation of the bound β -tubulin in β I differed from that of β II and β III. β I was substantially digested after only 5 min incubation with chymotrypsin, producing smaller size species. In contrast, β II and β III were substantially more resistant to proteolysis (>30% of initial protein was still present after 20 min). We conclude that release of β -tubulin from β I is associated with a conformational change from a protease-sensitive, 'unfolded' form to a relatively protease-resistant, assembly-competent, apparently folded structure.

These observations indicated that β I might represent a complex between β -tubulin polypeptides and a molecular chaperone. Because release of proteins from chaperone complexes requires hydrolysis of ATP^{26–29}, pulse-chase experiments were done in the presence of apyrase²⁹ (Fig. 2c). When apyrase was added just before chase, β -tubulin remained in the β I form

§ To whom correspondence should be addressed.

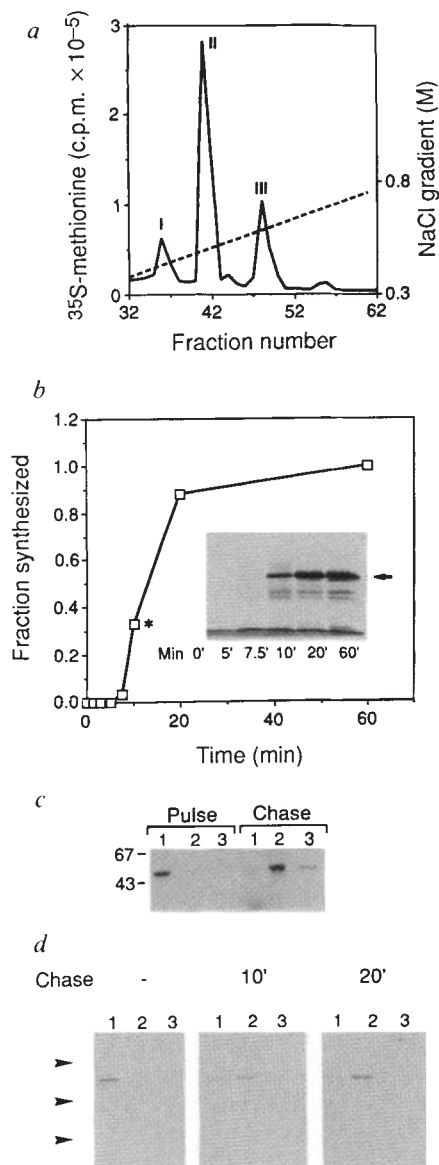


FIG. 1 β I is a precursor to the free β -subunit and heterodimer. *a*, Anion-exchange chromatogram of a 60-min translation reaction demonstrates multiple β -tubulin forms²¹. β I, β II and β III emerge as distinct peaks in fractions I, II and III, respectively. *b*, Pulse-radiolabelling of newly translated β -tubulin. Arrow in inset indicates full-length β -tubulin. Asterisk indicates 10-min time point. *c*, *d*, Pulse-chase analyses using 10-min synthesis reactions, and chase periods of 65 min (*c*), or 10 min and 20 min (*d*). Lanes 1, 2 and 3, Mono-Q peak fractions I, II and III, respectively. Arrows in *d* denote positions of the M_r standards (67, 43 and 29K). Conversion of β I into β II and β III is complete within 20 min at 30 °C. In contrast, at 4 °C no release of bound polypeptides from β I is detected over at least 8 h. **METHODS.** Synthetic messenger RNA for chicken β -tubulin was translated (30 °C) in the presence of ^{35}S -methionine in rabbit reticulocyte lysate (Promega) following ref. 21, and the radiolabelled β -tubulin products were analysed as a function of reaction time. In *a* the translation reaction was treated with DNase I and RNase A (10 $\mu\text{g ml}^{-1}$, 15 min, 30 °C), and eluted on a Mono-Q anion-exchange column with a linear gradient (dashed line) of 0–1 M NaCl in 20 mM sodium phosphate, pH 6.8, supplemented with 0.5 mM MgCl_2 and 20 μM GTP (flow rate 1 ml min⁻¹); 0.5 ml fractions were collected. *b*, Translation reactions were terminated with TS cocktail (final concentrations: 10 $\mu\text{g ml}^{-1}$ cycloheximide, 500 $\mu\text{g ml}^{-1}$ RNase A, 1 mM methionine) and analysed by SDS-PAGE fluorography. *c*, *d*, Translation was terminated with TS cocktail, and half (*c*) or one-third (*d*) of the sample was placed on ice ('pulse') while the remaining portions were incubated at 30 °C for the designated additional times ('chase'). The samples were then chromatographed as in *a* and fractions I–III were analysed by SDS-PAGE fluorography.

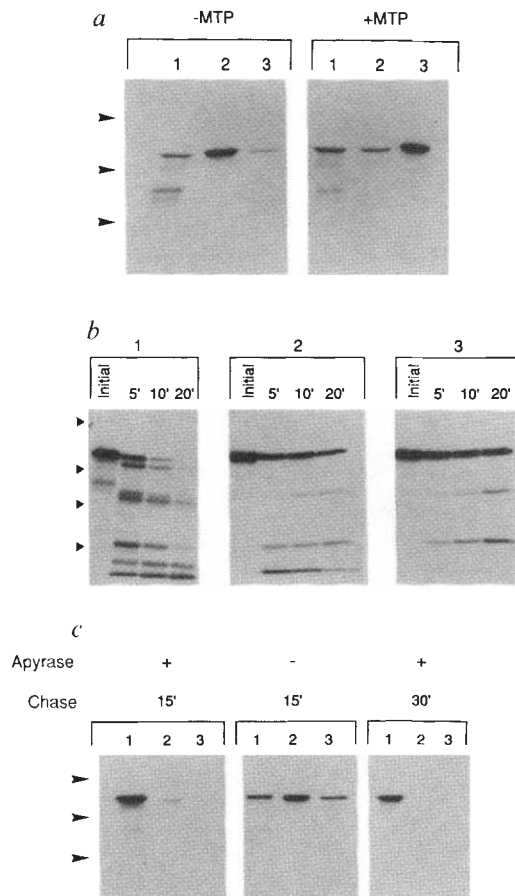


FIG. 2 ATP-dependent release of newly translated protein from β I yields monomers that are competent to form heterodimers and are relatively resistant to protease. *a*, β I chases into monomer in the absence of MTP and into heterodimer in the presence of added MTP. Lanes 1, 2 and 3, β I, β II and β III fractions. *b*, Limited proteolysis studies showing protease sensitivity of β I, β II and β III. *, Proteolytic bands in β I not observed or present at reduced intensity in β II or β III. *c*, Apyrase blocks release of β -tubulin from the β I form. Arrows denote positions of M_r standards (67, 43 and 29K in *a* and *c*; 67, 43, 29 and 20K in *b*). **METHODS.** *a*, A β -tubulin translation reaction (15 min synthesis) was terminated with TS cocktail and divided into two aliquots. One aliquot was given buffer only (–MTP); the other aliquot was brought to 1 mg ml⁻¹ MTP (+MTP). Heparin (50 $\mu\text{g ml}^{-1}$) was included to suppress microtubule polymerization²¹. Both aliquots were chased for 15 min at 30 °C and then chromatographed on the Mono-Q column and analysed as in Fig. 1. *b*, A 14-min translation reaction was terminated with TS cocktail and divided into three aliquots: one was placed on ice to arrest β -tubulin in the β I form; a second was chased for 30 min at 30 °C to convert β I into β II; and the third was supplemented with MTP (~1 mg ml⁻¹) and heparin (50 $\mu\text{g ml}^{-1}$) and chased for 30 min at 30 °C to convert β I into β III. The aliquots were chromatographed on the Mono-Q column. Peak fractions of β I, β II or β III were diluted 1:1 in PB+2.5Mg (pH 6.8) stabilizing buffer²², supplemented with BSA (2 mg ml⁻¹), digested at room temperature for the indicated times with chymotrypsin (Worthington; 8 $\mu\text{g ml}^{-1}$) and analysed by SDS-PAGE fluorography. *c*, A β -tubulin translation reaction (15 min synthesis) was divided into three aliquots. Two were supplemented with TS cocktail and apyrase (Sigma, 20 U ml⁻¹ final concentration) and chased at 30 °C for 15 or 30 min. Thin-layer chromatography confirmed that lysate ATP was rapidly hydrolysed to monophosphate by the apyrase (completed within ~5 min). The remaining aliquot (apyrase-free control) was supplemented with TS cocktail and chased for 15 min. The samples were chromatographed on the Mono-Q column and analysed as in Fig. 1.

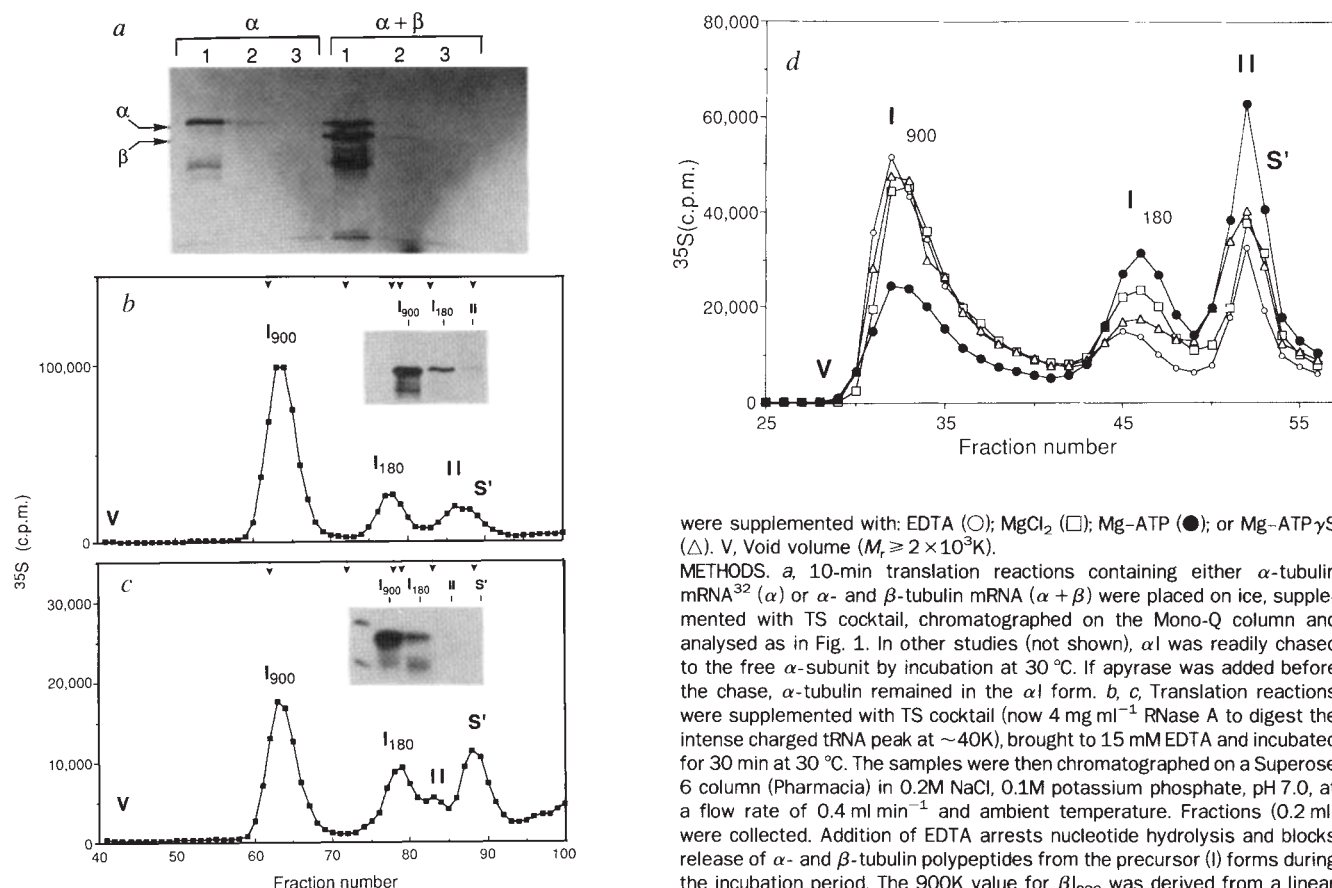


FIG. 3 Newly translated α - and β -tubulin enter a 900K complex. *a*, α -Tubulin enters the same early eluting complex as β -tubulin. *b*, *c*, Size-exclusion chromatogram of a 14-min translation reaction of β - and α -tubulin, respectively. Insets show SDS-PAGE fluorograms of the peak fractions. Lane 1 of the insets shows M_r markers (67 and 43K). S', The residual methionyl-transfer RNA complex. V, Void volume ($M_r \geq 3 \times 10^4$ K). Arrows indicate positions of molecular mass standards, left to right: thyroglobulin (669K), ferritin (440K), catalase (232K), aldolase (158K), bovine serum albumin (67K) and adenylate kinase (32K). *d*, Release of β -tubulin from the 900K complex requires Mg-ATP. Nucleotide-depleted 14-min translation reactions

were supplemented with: EDTA ($\circ$); MgCl_2 ($\square$); Mg-ATP ($\bullet$); or Mg-ATP γ S ($\triangle$). V, Void volume ($M_r \geq 2 \times 10^5$ K).

METHODS. *a*, 10-min translation reactions containing either α -tubulin mRNA³² (α) or α - and β -tubulin mRNA ($\alpha + \beta$) were placed on ice, supplemented with TS cocktail, chromatographed on the Mono-Q column and analysed as in Fig. 1. In other studies (not shown), α I was readily chased to the free α -subunit by incubation at 30 °C. If apyrase was added before the chase, α -tubulin remained in the α I form. *b*, *c*, Translation reactions were supplemented with TS cocktail (now 4 mg ml⁻¹ RNase A to digest the intense charged tRNA peak at ~40K), brought to 15 mM EDTA and incubated for 30 min at 30 °C. The samples were then chromatographed on a Superose 6 column (Pharmacia) in 0.2M NaCl, 0.1M potassium phosphate, pH 7.0, at a flow rate of 0.4 ml min⁻¹ and ambient temperature. Fractions (0.2 ml) were collected. Addition of EDTA arrests nucleotide hydrolysis and blocks release of α - and β -tubulin polypeptides from the precursor (I) forms during the incubation period. The 900K value for β I₉₀₀ was derived from a linear regression analysis of the semi-log plot of molecular mass versus elution position of the molecular mass standards, and may be a slight over- or underestimation of the true M_r value. *d*, A β -tubulin translation reaction (14 min synthesis) was supplemented with TS cocktail (4 mg ml⁻¹ RNase A) and centrifuged through five successive Sephadex G-25 spin columns in PB + 2.5Mg buffer to remove nucleotides and other low M_r material (5 ml beds; Isolab). The effluent was divided into four aliquots, supplemented with either EDTA (15 mM), MgCl_2 (2 mM), Mg-ATP (2 mM) or Mg-ATP γ S (2 mM), chased for 30 min at 30 °C and fractionated on a Superose 12 size-exclusion column (Pharmacia) using the phosphate/NaCl buffer above (flow rate of 0.5 ml min⁻¹ at ambient temperature). Fractions of 0.25 ml were collected.

('+apyrase'); when it was omitted, β I chased into β II and β III ('-apyrase'). The foregoing analyses were also done on α -tubulin with remarkably similar results (Fig. 3*a*). With both the α - and β -tubulin message, α I as well as β I were produced demonstrating that cotranslation does not alter the processing of the newly translated subunits (Fig. 3*a*).

β I generated at early times eluted on the Mono-Q column as an asymmetric peak at ~0.46 M NaCl. Analysis of this peak by size-exclusion chromatography demonstrated two size classes: a 140–180K form previously observed (β I₁₈₀)²¹, and β I₉₀₀, an undescribed form ($M_r \sim 900$ K) which was the major constituent. Given its abundance, β -tubulin associated with the 900K complex is apparently the precursor of the free β -subunit and heterodimer. When a translation reaction was directly subjected to size-exclusion chromatography, the major β -tubulin-containing product eluted as a 900K complex at the identical fraction position as the Mono-Q-purified complex (Fig. 3*b*). Figure 3*c* shows a similar result for α -tubulin. Release of β -tubulin monomer from the 900K complex was achieved in reconstitution experiments by the addition of Mg-ATP, further confirming that this complex was the precursor form (Fig. 3*d*). When Mg-ATP γ S was substituted for Mg-ATP, release of β -tubulin from β I₉₀₀ was negligible.

The 900K complex was purified from a β -tubulin translation

reaction by size-exclusion followed by anion-exchange chromatography, and subjected to SDS-polyacrylamide gel electrophoresis (PAGE) (Fig. 4). A prominent band(s) at 58K was detected together with several minor species of nearly the same apparent size. Considering recent suggestions that t-complex polypeptide-1 might function in the eukaryotic cytosol as a molecular chaperone¹⁻³, and noting that TCP1 resides in a high- M_r complex comprising a set of polypeptides of the same apparent 55–60K size³³, we tested whether the 58K species was in fact TCP1 by immunoblot analysis with a monoclonal antibody directed against the mouse protein³⁰. The 58K species demonstrated strong immunoreactivity (Fig. 4*d*). Furthermore, this species and the 900K complex displayed identical elution profiles (Fig. 4*a–c*). To establish that the immunoreactive 58K species was rabbit TCP1, we did an immunoblot analysis of a glycerol gradient fractionation of rabbit testis; a single reactive species from the 20S fraction exactly comigrated with the 58K component of the 900K reticulocyte complex (not shown). We conclude that the 900K complex from reticulocyte lysate is TCP1 complex.

These findings are consistent with a function for TCP1 complex as a molecular chaperone in tubulin biogenesis: (1) newly translated tubulin subunits, both α and β , are bound by this complex and released in assembly-competent forms; (2) release

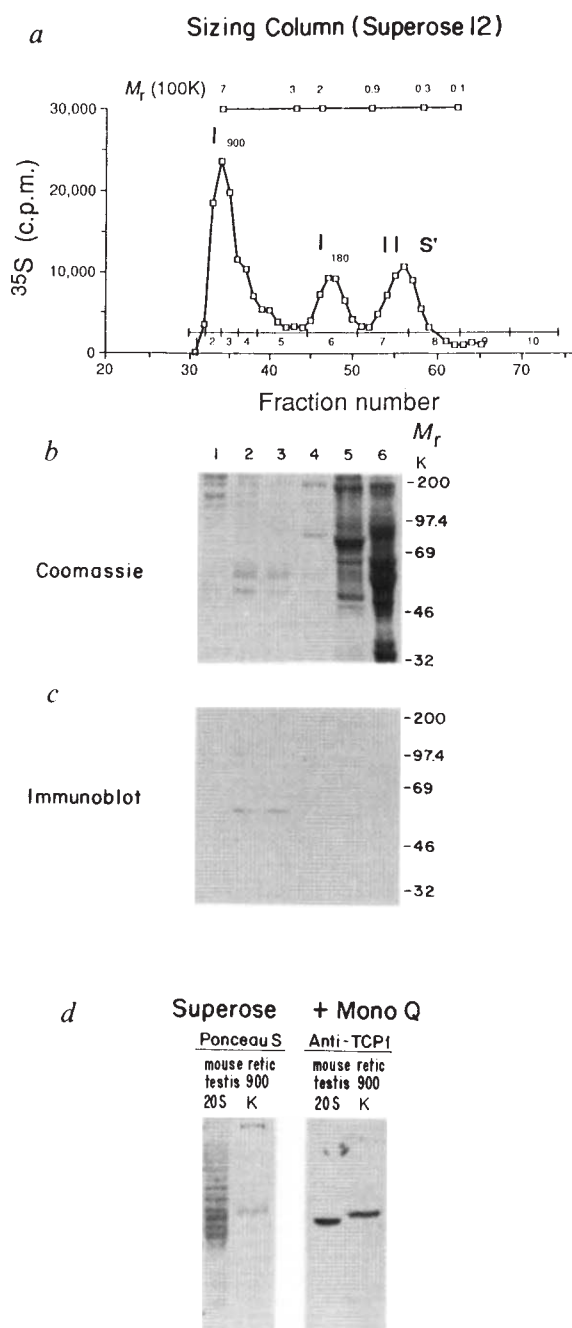


FIG. 4 The 900K complex contains a 58K species that cross-reacts with antibody to TCP1. *a-c*, Superose, β_{1900} copurifies with a 58K polypeptide that immunoreacts with a monoclonal antiserum against mouse TCP1³⁰ (antibody 91A) (provided by K. Willison). *d*, Superose + mono-Q, the 900K complex purified by size-exclusion and anion-exchange chromatography comprises a set of polypeptides (M_r ~55–60K), one of which reacts with antibody 91A.

METHODS. *a-c*, β -Tubulin translation reaction (300 μ l; 14 min synthesis) was supplemented with TS cocktail (4 mg ml⁻¹ RNase A) and 15 mM EDTA, incubated for 30 min at 30 °C and chromatographed on the Superose 12 column. Fractions (0.25 ml) were collected and pooled as indicated (*a*). The samples were dialysed against ammonium bicarbonate (1 mM), and concentrated to 50 μ l for SDS-PAGE. A 58K immunoreactive species was detected at the identical elution position as β_{1900} (immunoblot lanes 2 and 3 from pooled samples 2 and 3). Samples 7 and 8 ($10 \leq M_r \leq 90$ K) could not be analysed owing to a large excess of haemoglobin. The TCP1-reactive species also copurified with β_{1900} on the Mono-Q column, eluting at ~0.45 M NaCl (data not shown). *d*, A 1-ml β -tubulin translation reaction (14 min synthesis) treated as above was chromatographed on the Superose 6 column. Fractions 61–67 containing β_{1900} were applied to the Mono-Q column. Fractions 34–38 from the Mono Q column were collected, dialysed and concentrated (purified 900K complex). In addition to the 55–60K species, several slowly migrating species were sometimes observed as here, but not in all preparations. Mouse testes were homogenized (20 mM HEPES, 0.1% Triton-X100, pH 7.4), the extracts were centrifuged (14,000 r.p.m., 12 min at 4 °C), and the supernatants fractionated by sedimentation in a 10–30% (v/v) glycerol gradient (25,000 r.p.m., 24 h at 4 °C, Beckman SW41 rotor). The TCP1 complex was localized to the 20S fraction by immunoblot analyses with antibody 91A.

deficiency contribute to the observed phenotype remains to be determined. As suggested here, the TCP1 complex seems to play a role in the folding of tubulin. Whether it also plays a direct role in assembly of heterodimers or of microtubules themselves remains to be addressed. Preliminary studies suggest that TCP1 complex associates with several other newly translated proteins besides tubulin (unpublished data). This chaperone could therefore have a more general role in protein folding in the eukaryotic cytosol. □

Received 27 February; accepted 2 June 1992.

1. Ellis, R. J. *Science* **250**, 954–959 (1990).
2. Gupta, R. S. *Biochem. Int.* **20**, 833–841 (1990).
3. Trent, J. D., Nimmegern, E., Wall, J. S., Hartl, F.-U. & Horwich, A. L. *Nature* **354**, 490–493 (1991).
4. Dustin, P. *Microtubules* (Springer, New York, 1984).
5. Kirschner, M. & Mitchison, T. *Cell* **45**, 329–342 (1986).
6. Vallee, R. *Nature* **352**, 187–188 (1991).
7. Skoufias, D. A., Burgess, T. L. & Wilson, L. J. *Cell Biol.* **111**, 1929–1937 (1990).
8. Huffaker, T. C., Thomas, J. H. & Botstein, D. *J. Cell Biol.* **106**, 1997–2010 (1988).
9. Ahmad, S., Ahuja, R., Venner, T. J. & Gupta, R. S. *Molec. cell. Biol.* **10**, 5160–5165 (1990).
10. Bollag, D. M. *et al.* *Eur. J. Cell Biol.* **51**, 295–302 (1990).
11. Burke, D., Gasdaska, P. & Hartwell, L. *Molec. cell. Biol.* **9**, 1049–1059 (1989).
12. Caron, J. M., Jones, A. L., Rall, L. B. & Kirschner, M. W. *Nature* **317**, 648–651 (1985).
13. Cleveland, D. W., Kirschner, M. W. & Cowan, N. J. *Cell* **15**, 1021–1031 (1978).
14. Coias, R., Galego, L., Barahona, I. & Rodrigues-Pousada, C. *Eur. J. Biochem.* **175**, 467–474 (1988).
15. Isaacs, W. B. & Fulton, A. B. in *Cellular and Molecular Biology of Muscle Development* (eds Kedes, L. H. & Stockdale, F. E.) 137–146 (Liss, New York, 1989).
16. Redmond, T. *et al.* *Eur. J. Cell Biol.* **50**, 66–75 (1989).
17. Sanchez, E. R. *et al.* *Molec. Endocr.* **2**, 756–760 (1988).
18. Weinstein, B. & Solomon, F. *Molec. cell. Biol.* **10**, 5295–5304 (1990).
19. Wu, J. & Yarbrough, L. R. *Gene* **61**, 51–62 (1987).
20. Yen, T. J., Machlin, P. S. & Cleveland, D. W. *Nature* **334**, 580–585 (1988).
21. Yaffe, M. B., Farr, G. W. & Sternlicht, H. *J. Biol. Chem.* **263**, 16023–16031 (1988).
22. Yaffe, M. B., Farr, G. W. & Sternlicht, H. *J. Biol. Chem.* **264**, 19045–19051 (1989).
23. Grasso, J. A. *Anat. Rec.* **156**, 397–414 (1966).
24. Yaffe, M. B., Levinson, B. S., Szasz, J. & Sternlicht, H. *Biochemistry* **27**, 1869–1880 (1988).
25. Brown, H. R. & Erickson, H. P. *Archs biochem. Biophys.* **220**, 46–51 (1983).
26. Munro, S. & Pelham, H. R. *Cell* **46**, 291–300 (1986).
27. Rothman, J. E. *Cell* **59**, 591–601 (1989).
28. Gething, M.-J. & Sambrook, J. *Nature* **355**, 33–45 (1992).
29. Beckmann, R. P., Mizzen, L. A. & Welch, W. J. *Science* **248**, 850–854 (1990).
30. Willison, K. *et al.* *Cell* **57**, 621–632 (1989).
31. Ursic, D. & Culbertson, M. R. *Molec. cell. Biol.* **11**, 2629–2640 (1991).
32. Cowan, N. J., Dobner, P., Fuchs, E. V. & Cleveland, D. W. *Molec. cell. Biol.* **3**, 1738–1745 (1983).
33. Lewis, V. A., Hynes, G. M., Zheng, D., Saibil, H. & Willison, K. *Nature* **358**, 249–252 (1992).

ACKNOWLEDGEMENTS. We thank K. Willison and J. Trent for biological materials and W. J. Welch for discussions. This work was supported by the American Cancer Society and the Howard Hughes Medical Institute.

of the subunits requires hydrolysis of ATP; (3) conformation of the tubulin subunits is altered as the result of interaction with the complex (from a protease-sensitive, apparently unfolded form to a protease-resistant, native-like form), implying a role in folding of newly translated tubulin subunits. It thus seems likely that the hydrolysis of ATP required for release of tubulin subunits from the complex is in fact associated with folding to a native form.

Our conclusions about the role of TCP1 in tubulin biogenesis seem to extend to the intact cell. In CHO cells we observed similar associations of newly translated tubulin with a 900K complex (H.S. *et al.*, manuscript in preparation). Our finding that the TCP1 complex is involved in tubulin biogenesis is not entirely unexpected, given the previous observation that mitotic spindle assembly is impaired in a cold-sensitive TCP1 mutant of *Saccharomyces cerevisiae*³¹. Whether the spindle defects observed in the mutant can be explained entirely by a defect of microtubule biogenesis or whether other effects of TCP1

T-complex polypeptide-1 is a subunit of a heteromeric particle in the eukaryotic cytosol

Victoria A. Lewis, Gillian M. Hynes, Dong Zheng*, Helen Saibil* & Keith Willison

Institute of Cancer Research, Chester Beatty Laboratories, Fulham Road, London SW3 6JB, UK

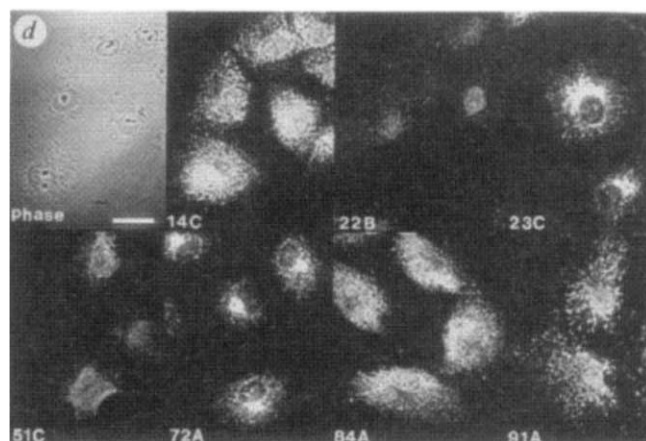
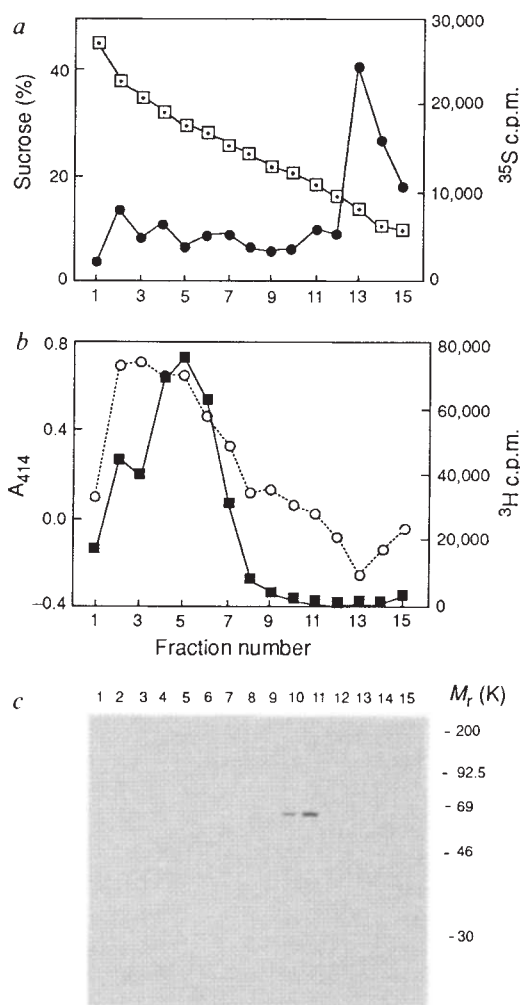
* Department of Crystallography, Birkbeck College, London WC1E 7HX, UK

THE murine t-complex encodes t-complex polypeptide-1 (TCP1)¹, which is constitutively expressed in almost all cells, and upregulated during spermatogenesis^{2,3}. Mammalian sequences have >96% identity with each other, and >60% identity with *Drosophila melanogaster* and yeast orthologues⁴⁻⁹. TCP1 is essential in yeast⁹, and is postulated to be the cytosolic mammalian equivalent of groEL^{10,11}. We report here that, in the native state, murine and human TCP1 is distributed throughout the cytosol as an 800K–950K hetero-oligomeric particle in association with four to six unidentified proteins and two Hsp70 heat-shock proteins. Negative-stain electron microscopy indicates that the structure is two stacked rings, 12–16 nm in diameter. Therefore, despite similarities with the chaperonin 60 proteins, these data indicate that TCP1 is biochemically and structurally unique. We suggest that TCP1 may represent one of a family of molecules in the eukaryotic cytosol involved in protein folding and regulated in part by their heteromeric associations.

The subcellular distribution of TCP1 was determined using sucrose-gradient fractionation of HEp-2 cell homogenates (Fig. 1). Analysis of the gradient fractions revealed that TCP1 did not sediment with the intact membranous organelles or as a monomer of M_r 60K. Instead, TCP1 banded in fractions 9 and 10, corresponding to 19.5–21.5% sucrose (1.0783–1.0877 g cm⁻³) and an apparent M_r of 800K–950K, although these fractions contained only 0.02–0.07% of the total cell protein. The sedimentation of TCP1 was reproducible on identical gradients

FIG. 1 *a, b, c*, Fractionation of HEp-2 cell post-nuclear supernatant (PNS) homogenates on 10.2–40% linear sucrose gradients. *a*, Sucrose densities are displayed as per cent sucrose per fraction (open squares). The distribution of total protein per fraction is indicated by incorporated ³⁵S-methionine c.p.m. (hatched ovals). The bulk of the cellular protein was recovered as soluble material at the low density end (top) of the gradient. *b*, The majority of the Golgi and lysosomal membranes remained intact during the procedure and were recovered in fractions 2–7 (open circles (³H, c.p.m.) and hatched squares (absorbance at 414 nm, A_{414}), respectively), corresponding to a sucrose density of 1.1031–1.1612 g cm⁻³ (24.9–37.0%). *c*, The sedimentation of TCP1 at 19.5–21.5% sucrose was determined by western blot analysis of gradient fractions probed with the monospecific rat anti-TCP1 monoclonal antibody 84A (ref. 2). *d*, Indirect immunofluorescence micrographs of CV-1 cells stained with individual rat anti-TCP1 monoclonal antibodies designated 14C, 22B, 23C, 51C, 72A, 84A and 91A, respectively², and detected by rabbit anti-rat-conjugated FITC. Phase-contrast image of staining with 14C is shown immediately to the left. Monkey cells were used because western blotting had shown that antibodies 14C, 84A and 91A were monospecific for TCP1 in CV-1 cells, and antibody 22B acted as an internal negative control, failing to detect any monkey antigens (V.A.L. *et al.*, manuscript in preparation). Antibodies 23C, 51C and 72A were not monospecific in CV-1 cells and failed to give a similar diffuse cytoplasmic distribution.

METHODS. HEp-2 cells (1.0×10^8) were collected, washed once with ice-cold Harms buffer (250 mM sucrose, 10 mM triethanolamine, 10 mM acetic acid, 1 mM EDTA, pH 7.5), repelleted, resuspended into Harms buffer and left on ice for 10 min. Cells were homogenized using a steel ball-bearing homogenizer, and the PNS was recovered by centrifugation (3,000 r.p.m., 10 min, 4 °C). Protease inhibitors (0.3 U ml⁻¹ aprotinin, 1 mM PMSF, 5 mg ml⁻¹ chymostatin, 5 mg ml⁻¹ pepstatin A, 5 mg ml⁻¹ antipain, 10 mg ml⁻¹ leupeptin) were added before layering the PNS onto prechilled (0 °C) continuous sucrose gradients (10.2–40% (w/w) in 90 mM KCl, 50 mM



HEPES pH 7.2, plus protease inhibitors). Gradients were centrifuged in a Beckman SW-28 rotor (25,000 r.p.m. 16 hours, 4 °C) and fractionated into 15 fractions of 1.1 ml from bottom (fraction 1) to top (fraction 15). Apparent molecular mass of the TCP1 heteropolymer was extrapolated from sedimentation of known native proteins (thyroglobulin (669K), Jack bean urease (545K), apoferritin (443K), β -amylase (200K), bovine serum albumin (69K) and ovalbumin (45K)) run on a parallel gradient. The sedimentation of Golgi membranes was determined by a galactosyl transferase assay²²; and of lysosomes by a β -hexose-aminidase assay. Western blot analysis and SDS-PAGE were as described² except permeabilization was by 5 min incubation in 100% MeOH at –20 °C. All secondary antibodies were obtained from Pierce.

FIG. 2 Purification of TCP1 as a protein complex by a two-step method of sucrose gradient fractionation followed by either DE-52-cellulose anion-exchange or ATP affinity chromatography. **a**, Silver staining of SDS-PAGE-separated samples from the purification of murine testis germ cell TCP1 complex: PNS (lane 1); pooled 19.5–21.5% sucrose gradient fractions (lane 2); TCP1 peak eluates after DE-52 and ATP-agarose chromatography (lanes 3 and 4, respectively). Samples obtained from ATP-agarose chromatography (lane 5) or immunoprecipitation with anti-TCP1 antibody 91A (lane 6) were western blotted and probed with monoclonal mouse anti-Hsp70. 35 S-methionine-labelled proteins with BHK or murine testis germ cells were immunoprecipitated with anti-TCP1 antibody 91A, separated by SDS-PAGE and autoradiographed as shown in lanes 7 and 8, respectively. Immunoprecipitation in mixed micelle buffer of the same germ cell preparation used in lane 8 with anti-TCP1 antibody 91A identifies TCP1 in the complex (lane 9). **b**, **c**, **d**, Hep-2 cell (**b**) or germ cell (**c**, **d**) TCP1 complex purified by ATP-agarose chromatography, separated by two-dimensional isoelectric focusing SDS-PAGE and silver stained (**b**, **c**) or transferred to nitrocellulose and probed with anti-TCP1 antibody 91A to identify TCP1. **d** The Hsp70 proteins are indicated by arrowheads and TCP1 by an arrow.

METHODS. Either $1.8\text{--}2.0 \times 10^7$ Hep-2 cells, or $5.0\text{--}8.0 \times 10^8$ testis germ cells (prepared as described³) were used for one round of purification. Sucrose gradient centrifugation was as for Fig. 1. The two fractions corresponding to 19.5–21.5% sucrose were pooled and dialysed against column equilibration buffer (25 mM Tris-acetate pH 7.4, 0.1% TX-100, 20 mM NaCl, 5 mM MgCl₂) at 4 °C, then loaded onto ATP-agarose (Sigma) or DE-52-cellulose (Whatman) columns (4 mm $\times$ 200 mm) equilibrated at 4 °C. Columns were eluted with linear gradients of 0–15 mM ATP or 0–0.4 M NaCl, respectively, in 5 bed-volumes of equilibration buffer. The eluate was analysed on DOT-BLOTs to determine the peak fractions of TCP1 which was seen at 60 mM NaCl (DE-52) and 3 mM ATP (ATP-agarose). The pooled peak was concentrated using an Amicon Centricon 30 microconcentrator to a final volume of 200 μ l, rinsed once with 25 mM Tris-acetate; 5 mM MgCl₂;

5 mM ATP and reconcentrated to 200 μ l. Each purification from 1.0×10^9 germ cells resulted in reproducible yields of about 300 μ g, and from tissue culture cells 60–100 μ g per 2.0×10^8 cells. Recovered samples were used immediately for further analysis or snap-frozen in liquid nitrogen and stored at -70 °C. Recovery of TCP1 was assessed at all stages by western blot analysis and silver-stained samples separated by 10% SDS-PAGE as described². Monoclonal mouse anti-Hsp70 antibody (IgG₁; clone 3a3; Affinity Bioreagents, New Jersey, USA) and used at a dilution of 1:5,000. Two-dimensional PAGE was a modification of the O'Farrell's procedure as described by Corbett and Dunn²³. IEF gels were run for 12,000 volt-hours producing a linear gradient of pH 4.5–7.5. The second dimension SDS-PAGE gels were 8% acrylamide.

of mouse testis germ cells, BHK, NRK, CHO, CEM, and *Schizosaccharomyces pombe* cell homogenates. These data suggest that native TCP1 exists as a high- M_r polypeptide complex common to a broad range of species and cell types.

Contrary to previous immunolocalization using rat-anti-TCP1 monoclonal antibodies², TCP1 fractionated away from the membrane components, therefore we reassessed the subcellular distribution of TCP1 by indirect immunofluorescence. Three anti-TCP1 monospecific antibodies (Fig. 1d; 14C, 84A, 91A) gave a diffuse cytoplasmic staining pattern. The cytoplasmic localization was reproducibly observed in different cell lines and, taken together with the fractionation data, demonstrated that TCP1 existed in a non-membrane-bound form. The perinuclear staining pattern previously identified as specific to membranes of the *trans*-Golgi network² was only found in

cells stained with antibodies 23C and 72A (Fig. 1d). Western blotting has shown that, despite being monoclonal, neither 23C nor 72A was monospecific for TCP1, and that a 92K protein was the antigen responsible for the previous localization to the *trans*-Golgi network (V.A.L., unpublished results).

TCP1 was purified from the post-nuclear supernatant of Hep-2 or mouse germ-cell homogenates by a two-step method (Fig. 2). The first step involved linear sucrose gradients as shown in Fig. 1 because >95% of the TCP1 was recovered in two fractions, representing an enrichment of about 500-fold. SDS-PAGE analysis of the material recovered from the second step (Fig. 2a: DE-52 or ATP-agarose column chromatography, lanes 3 and 4, respectively), revealed eight polypeptides from germ-cell post-nuclear supernatant in addition to TCP1 with M_r ranging from 72K to 53K. Two-dimensional SDS-PAGE (Fig.

FIG. 3 Electron micrographs of preparations of purified murine testis germ cell TCP1 protein complex. Lower magnification views (**a**, $\times 100,000$) indicated the relative uniformity of the preparation. Higher magnification of representative individual structures (**b**–**e**, $330,000\times$) display images of both ring-like end views and rectangular side-views.

METHODS. Solutions of purified TCP1 complex ($\sim 0.1\text{--}0.2 \mu\text{g } \mu\text{l}^{-1}$) were negatively stained on thin carbon support films with either 2% sodium phosphotungstate at pH 7 or 2% uranyl acetate. Electron micrographs were recorded at 120 kV using a JEOL 1200EX microscope.

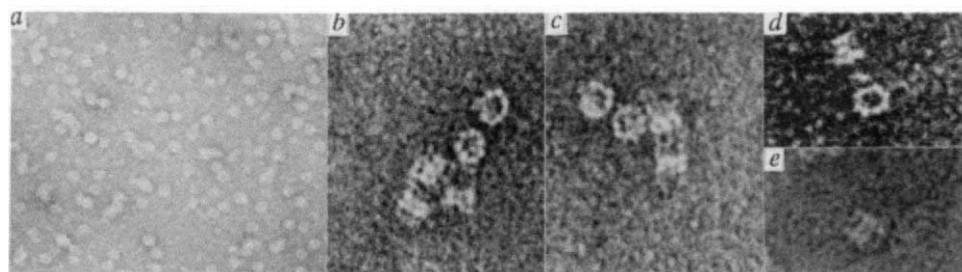


FIG. 4 Alignment and comparison of the deduced amino-acid sequence for murine TCP1B polypeptide (EMBL P11983) with human TCP1 (TCP1Hu; EMBL P17987)⁴, *Drosophila melanogaster* TCP1 (TCP1Dm; EMBL P12613)⁸, and *Saccharomyces cerevisiae* TCP1 (TCP1Sc; EMBL P12612)⁹; TF-55 from *Sulfolobus shibatae*¹⁷; and the following cpn60 members: *Mycobacterium leprae* 65K antigen (65K, EMBL P09239)²⁴, Rubisco binding-protein alpha subunit from *Triticum aestivum* (RuBP; EMBL P08823)²⁵, GroEL (groEL; EMBL P06139)²⁵, and human Hsp60 (EMBL P10809; R. S. Gupta, unpublished results). Arrangement of sequences from top to bottom reflects the corresponding degree of homology to TCP1B. Amino-acid numbers are relative to the murine TCP1B sequence with the required gap insertions (—) to maximize the number of matches and alignment. The full amino-acid number is given at the end of each sequence. Where amino acids 5' to the start are omitted, the first residue number is given at the beginning of the sequence. Matches are indicated by upper case, and mismatches by lower case letters. Regions of high homology are bold.

METHODS. Homologies with TCP1 were identified using the Protein search program from J. F. Collins and A. F. W. Coulson of the Bio-computing Research Unit in Molecular Biology, University of Edinburgh, Scotland, which recognizes short stretches of conserved primary and secondary structure and takes into account conserved amino acid changes. Alignment of the nine polypeptide sequences was assisted by the MacVector program (IBI Ltd) for the Apple Macintosh.

	10	20	30	40	50	60	70	80
TCP1b	MEGPLSVFGD	RSTGEAVRSQ	NVMAAASIAN	IVKSSFGPGV	LDKMLVDDIG	DVITITNDGAT	ILKLEVEHP	AAKVLG----
TCP1Hu	MEGPLSVFGD	RSTGETIRSQ	NVMAAASIAN	IVKSSIGPGV	LDKMLVDDIG	DVITITNDGAT	ILKLEVEHP	AAKVLG----
TCP1Dm	(4) lasPLSlaGt	RsrriRVRQ	NVMAALSIAN	IVKSSIGPGV	LDKMLVDDIG	DVITITNDGAT	ILRLLEVEHP	AAKVLV----
TCP1Sc	(11) rsdtLflgGe	klsGGdiRQ	NVLATmavAN	vKSSIGPGV	LDKMLVDDIG	DfTITNDGAT	ILSLLDVqHP	AgKILV----
TF-55	(14) pvilIkeqss	RtyGkealra	NiaAvkaIee	alkStyGPG	mDKMfVDSIG	DiTITNDGAT	ILDkmdlqHP	tgKILV----
65K	makt	iaydEeaRrg	lgrlgnSLad	avKvtIGPG	rnvvLekkwG	apTITNDGvs	TeKeIEIEEdP	yeKigaelvk
RuBP	gadake	iafdqksRaa	lqagveklAN	avGvtIGPG	-rnvvLDeyG	nphkvNDGvT	IaraIElanP	menagaalir
groEL	maakd	vkfGndaRvk	mlrgvnvLad	avKvtIGPG	rnvvLDksfG	apTITkDGVs	vareIEIEEdk	femmgagmvk
hsp60	(17) lrraysshke	lkfGvegRas	llkgvetLae	avAatIGPG	rnvlieqfG	pphITkDGVt	vaKsivlkdk	femmgakllq
	90	100	110	120	130	140	150	160
	ELADLDQKEV	GDGTTSVVII	AEELLNADE	LVKQKIHPIS	VISGYRLACK	EAVRYINENL	IINTDELGRD	CLINTAKTSM
	ELADLDQKEV	GDGTTSVVII	AEELLNADE	LVKQKIHPIS	VISGYRLACK	EAVRYINENL	IvNTDELGRD	CLINaAKTSM
	ELAQLDQEV	GDGTTSVVII	AEELLNADE	LVKQKIHPIS	iISGYRIACK	EackYISEnL	tapvDELGRD	sLINIAKTSM
	ELAQqQDrEi	GDGTTSVVII	AsELLKAnE	LVKnKIHPIT	iItGfRvAlR	EaIRfINEvL	stsvDtLGke	tLINIAKTSM
	qIAqQDeEt	GDGTTaVil	AgELaKAd	LlykeIHPTI	ivsGYkkAee	iAlktIqG-i	atqpsvIndtD	vLrkvAlTSL
	EvAkktdDva	GDGTTatvl	AqaLvK--eg	L-r-nvaaga	nplIGkrgle	kAVdkvtEdT	lkdkdEvtek	eqI-aAtaai
	EvAkktdnsa	GDGTTacvL	ArEiiKigil	sVtsganPvS	lkkGidktvg	gileel-Erk	arvpkvsGdi	kavasiasgn
	EvAskandaa	GDGTTatvl	Aqaiiteglk	aVaagmnRmd	lkrGidkAvt	aAVEelk-aL	svpcsdskai	aqvgTisans
	EvAskntnea	GDGTTatvl	graiftesvk	nVaagcnRmd	lrrGsqvAv-	EkViefIsan	kkeittseei	aqvaTisang
	170	180	190	200	210	220	230	240
	SSKIINGINGD	YFANMVDAV	LAVKYTDARG	QPRY-PVNSV	NILKAHGRSQ	IESMLINGYA	LNCVVGSGQM	P-RRIV--NA
	SSKIINGINGD	ffANMVDAV	LAIKYTDIRG	QPRY-PVNSV	NILKAHGRSQ	mESMLISGYA	LNCVVGSGQM	P-RRIV--NA
	SSKIIGadade	ffsaMVDAa	qsVKITDPRG	QavY-sikaV	NvLKAHGKsr	rESvLIpGYA	LNCtiaSQqM	P-MRIV--NA
	SSKIIGadsd	ffsNMVDAI	LAVKtqnsqG	eikY-PVkaV	NvLKAHGKsa	tESILvpGYA	LNCtVaSQaM	P-MRIaggNv
	gSKavagare	YIAdLVKAV	aqV--aelRG	dkwYvldnV	qIVkKHGGSi	ndtqLVyGiv	vdkeVvhpGM	P-MRIE--NA
	Sagdqst-GD	lIAe-amDkV	gnevITvee	sntf-glqlE	ltegmrfdkg	yiSgyfvtDA	erqeaVleep	yillvsskvs
	delIGamiaD	aidkvpgDgV	LsiessssfE	ttvd-veegm	eIdrgyispQ	fvtnLeksiv	efenarvLit	dqK-Itsike
	detvgkliaa	amdkvgkegV	itVedgtqlq	deld-vVegm	qfdrgylsPy	finkpetGav	elnesopfill	adKISnIre
	dShvgkllas	amekvgekvG	itiregrtle	dele-vtegm	rfdrgfispY	fitdpksskv	efek-plill	seKLIssiqd
	250	260	270	280	290	300	310	320
	----KIACLD	FSLQKTKMKL	GVQVVITDPE	KLDQIRQRES	DITKERIQ-K	ILATGANVIL	-----TTG	GIDDMYLKYF
	----KIACLD	FSLQKTKMKL	GVQVVITDPE	KLDQIRQRES	DITKERIQ-K	ILATGANVIL	-----TTG	GIDDMcLKYP
	----KIACld	FSLQKTKMKM	GVQVlInDpd	KLeaIRaREL	DITKERIn-m	ILGTGVNVVL	-----vG	GvDdlcmKYF
	----KIACld	lnLQKArMam	GVQInIdDPE	qLEqIRkREA	qIVlERvK-K	IIdaGqGVVl	-----Ttk	GIDDLcKkEf
	----KIALLD	aSLeveKeP	dveirInDPT	gmhkfleeEe	nILKEkvD-K	IAATGANVIL	-----cqk	GDevaqHYL
	tvkdllpLe	kviQagKsLl	ii-aedvegE	aLstlvvnki	rgTfksvavK	apgfGdrrka	mlqdmallTC	aqviseevgl
	----iIplLe	qttQlrcplf	iV-aeditgE	aLatlvvnkl	rglinvaalK	apsfGerrka	vlqdiaVlTC	aeYlakdlgl
	----mlpVLe	avakagKpLl	ii-aedvegE	aLatavvnti	rglvkvaavK	apgfGdrrka	mlqdiatlTC	Gtvisseelgm
	----ilpaLe	iSnQsrpLl	ii-aedvdgE	aLaacilnkl	rggvkvcaVvK	apgfGdNrkN	tigdiavlTC	Gtvtfeeldl
	330	340	350	360	370	380	390	400
	VEAGAMAVRR	VLKRDILKHA	KASGASILST	LANLECEZTF	EVTMLGQAE	VVQERICDDE	LILIKN---T	KARTSASIIIL
	VEAGAMAVRR	VLKRDILKria	KASGATILST	LANLECEZTF	EaaMLGQAE	VVQERICDDE	LILIKN---T	KARTSASIIIL
	VEAGAMAVRR	VkKsDLKila	KATGaafits	LtNmDCERSF	dasMvGeAae	VaQERICDDE	LILIKG---T	KARaaASIIIL
	VEAKIMGVR	ckKedLrria	rATGaLvsS	msNLECEZTF	EssyLGldce	VWqakfsmDDE	CILIKG---T	akhsSsIIIL
	akkGILAVRR	akKsDLekla	rATGgrvIsn	ideLtsqdlg	yaa-LveerG	VkedKs----	-vfvveg---	-KnkpSvSIIl
	tlentdlS--	lLgkarKvVm	tkdettIV--	--egaGtda	iagrvaQirt	elensdsDyd	-reklqerla	KlaggvavIk
	lvenA-tVdq	-LgtarK--i	tlhqt--ttT	LIadaaakde	iqarvaQlkk	elsEatsDyd	-seklaria	KlsggvavIk
	elekAt----	-lledLgqak	rVvinkdttT	lIdgvCEaa	iqgrvaQlrrq	qEeatsDyd	-reklqerla	KlaggvavIk
	kpeqcti---	---enLgscd	sitvtkedvT	ilNgsQpkea	iqerieQikg	siditttnsy	ekeklqerla	KlsggvavIr
	410	420	430	440	450	460	470	480
	RGA-NDFMCD	EME-RSLHDA	LCVVKRVLEs	KSVVPGGAV	EAALSITYLEN	YATNMGSRQ	LAIAEFARSL	LVPINTLAVN
	RGA-NDFMCD	EME-RSLHDA	LCVVKRVLEs	KSVVPGGAV	EAALSITYLEN	YATNMGSRQ	LAIAEFARSL	LVPINTLAVN
	RGA-NDFYCD	EME-RSLHDA	LCVVKRVLEs	KkVvPGGcV	EAALSITYLEN	fATslasREQ	LAIAEFakSL	LVPKTLsVN
	RGA-Nbysld	EME-RSLHda	LsVVKRtLEs	gnVVPGGcV	EAALnIYLDN	fATtVGSREQ	LAIAEFaaal	LIIPKTLAVN
	Rgq-lervvd	E-E-RaLrDA	LgtVadVird	graVaGGGAV	EieiakrLrk	YApqvGgkEQ	LAleayAnaI	eglimiLaen
	aGaatevelk	Ekhkr-lEda	vrnaKaavEe	gi-VaGGGvt	llqaapaLdk	lkit-G-dEa	tganiVkvAl	-----
	vGaatevele	drqir-lEda	knaatfaaEe	gi-VPGGGAA	yvhLStYypa	iketiedhed	rlgAdiikq-	-----
	vGaatevele	EkaR-vEDA	LhatraavEe	g-VaGGGva	lirvaskLad	lrqg-nedqn	vgIKvalRam	eaplrqIvIN
	vGgasevevg	Ekkdr-ydDA	LhatraavEe	gi-lPGGGta	lvkaSrvLde	vvvd-nfdqk	IgvdiIrkal	trpakIieN
	490	500	510	520	530	540	550	560
	AAQDSTDLVA	KLRAFHNBAQ	-VNPE---RK	NLKW-IGLDL	VHGKPRDNKQ	AGVFETIVK	VKSIFATEA	AITILRIDDL
	AAQDSTDLVA	KLRAFHNBAQ	-VNPE---RK	NLKW-IGLDL	snGKPRDNKQ	AGVFETIVK	VKSIFATEA	AITILRIDDL
	AAKDaTDLVA	KLRsyHNssQ	-tkPE---Rs	dLKW-tGLDL	ieGvvrDNKK	AGVLEPamK	IKSLKFATEA	AITILRIDDM
	AAKDSseLVA	KLRsyHaasQ	makPEdvkRr	syzn-yGLDL	irGKivDeIh	AGVLEPTIak	VKSILKALEA	cvaILRIDtm
	AqldpDkLm	qLrSLHen--	-----e	tnKW-yGLnL	ftGPeDmWk	lGVIEPalVK	mnaKaATEA	vtvlvRIDDI
	-----eapl	KqIAFnsqme	pgvvaekVR-	NLsvghGLna	atGeyedLlk	AGvadPvkVt	sqalqnAasi	AgflfLteav
	AlQapasLiA	nnagvvegEv-	-----viekI	eseWemGyna	mtdKyenlie	sgVidPakVt	rcaIqnAasv	sgmvLttqai
	cgeepsvvan	tvkggdgnyg	ynaateeygn	m-----	-----id	mGldlPtkVt	rsaIqyAasv	AglmittecM
	Ageegsvilg	KlIdeygdff	agkydaskse	y-----	-----tDmla	tGldlPfkVv	rsqIwdAsqv	Asllatteva
	570	580						
	IKLHPE-SKD	DKHGSYENAV	HSGALDD	(556)				
	IKLHPE-SKD	DKHGSYEDAV	HSGALNd.	(556)				
	IKLHPE-dks	gK--SYadAc	aaGeLDg.	(558)				
	ItvdPEppKe	DpHdh.	(560)					
	vaagkk-ggs	epgGkEkee	kSsed.	(552)				
	vadkPE--Kt	aapaS-dptg	gmGmDf.	(541)				
	vvekPE-pKp	kvaepaE---	--GqLsv.	(543)				
	vtdlPEndaa	Dl-Gaaggmg	gmGmgmm.	(547)				
	IvdaPE--pp	aaaGagmpg	gmppgmppm.	(572)				

2b, c) resolved the individual TCP1 complex components and demonstrated that the mouse germ-cell proteins were strikingly similar to those from Hep-2 cells (Fig. 2b and c, respectively), with the exception that the Hep-2 cell oligomer contained two less proteins. Immunoprecipitation with anti-TCP1 antibodies under monospecific conditions yielded a group of proteins with similar SDS-PAGE mobilities to those from purified material (Fig. 2a, lanes 7 and 8), with the exception that the 72K and 70K species were present in lower amounts. Therefore, three independent purification methods (immunoprecipitation, DE-52 anion-exchange or ATP-agarose affinity chromatography) yielded protein complexes with components of similar apparent M_r and pI from species as diverse as human and mouse. Furthermore, pulse-chase experiments have shown that after a 5 min pulse, all the components of the TCP1 complex were co-immunoprecipitated within a 5 min chase, suggesting that TCP1 and its associated components were synthesized at similar rates and assembled within 10 min (V.A.L., unpublished results). These data strongly suggest that the TCP1 complex purified from mouse testis and Hep-2 cells was composed of TCP1 and eight or six other proteins, respectively, and that the heteromeric association was responsible for the high molecular mass of native TCP1. We propose that the proteins identified as copurifying with TCP1 be referred to as TAPs, or TCP1-associated polypeptides.

Owing to the discrepancy between the levels of 72K and 70K polypeptides recovered by co-immunoprecipitation versus purification, and because of their molecular masses and isoelectric points, we assessed their relationship to the Hsp70 family. Both polypeptides were recognized by an anti-Hsp70 antibody in purified TCP1 preparations and anti-TCP1 antibody immunoprecipitates (Fig. 2a, lanes 5 and 6, respectively) indicating that these proteins are indeed Hsp70s. The differential copurification of Hsp70s with the TCP1 complex involved an interaction sensitive to immunoprecipitation, but stable to the purification conditions which may reflect a weak, functionally regulated, association distinct from the association with the other TAPs. But binding to the DE-52 or ATP-agarose column may have resulted, at least in part, from a direct interaction with the Hsp70s in addition to TCP1 or the TAPs because the Hsp70 family is known to have a high affinity for ATP, and several members have previously been purified by a combination of DE-52 and ATP-affinity chromatography^{12,13}. We believe the native TCP1 complex does associate with the Hsp70s because they are consistently observed with the immunoprecipitated particle as well as the peak fractions eluted from DE-52 and ATP-agarose.

The purified TCP1 complex was visualized by negative stain electron microscopy revealing ring-like profiles enclosing a central channel with outer diameters ranging from 12 to 16 nm (Fig. 3). The images did not allow an exact determination of the radial symmetry, however, 8 or 9 subunits were consistently observed. Other views appeared as rectangular striated struc-

tures. The combination of ring and rectangular striated structures was reminiscent of negative stain images of cpn60 (refs 14, 15) suggesting a similar structure of two stacked rings. But the TCP1 particle lacked the groEL 7-fold symmetry and the individual subunits appeared smaller relative to the central channel, and less homogeneous in size and shape. Recently, an archaeobacterial cytoplasmic chaperone, TF-55, has been identified, similar to the *Pyrodictium* heat-shock inducible ATPase¹⁶, which displays a high degree of amino-acid homology with TCP1 (36–40% identity, see Fig. 4)¹⁷. The morphology of TCP1 most closely resembled the structures of TF-55, the *Pyrodictium* ATPase, and the functionally unrelated multicatalytic proteinase¹⁸ or prosome.

Alignment of the homologies between the four published TCP1 sequences, TF-55, and four members of the cpn60 family is shown in Fig. 4. The alignment shown is different from that previously published¹⁰ and identifies several novel stretches of between 5 and 10 amino acids conserved across all nine sequences. The most striking is a putative ATP-binding domain closely related to the cyclicAMP-dependent protein kinase¹⁹. The motif centres around the absolutely conserved GDGTT (single-letter amino-acid code) sequence at position 87–91 in TCP1B, however conserved changes extend the homology on either side. Although the overall homology is statistically weak, we believe the domains identified may prove to be biochemically and functionally significant.

On the basis of the limited sequence homology, it is suggested that TCP1 may be the mammalian cytosolic groEL equivalent^{10,11}. Along with the homology, both TCP1 and the chaperonins are abundant cytosolic proteins which form oligomeric structures of similar molecular mass. However, significant differences exist between TCP1 and the cpn60 members which indicate that, should TCP1 prove to be a functional chaperone²¹, it will not be a groEL equivalent. The expression of chaperonins increases as a result of stresses which, in turn, renders them potent antigens²⁰. In contrast, levels of TCP1 were unaffected by treatment with heat shock (V.A.L., unpublished results) and the production of anti-TCP1 antibodies has been difficult (ref. 2 and K.W., unpublished results). In striking contrast to the chaperonins, the TCP1 complex did not exist as a homopolymer, but as a heteromeric protein assembly containing at least seven different proteins. Several TCP1-related genes in humans (K.W., unpublished results and A. Malik, unpublished results), mice (A. Ashworth, unpublished results), yeast (B. Barrell, unpublished results), and *Caenorhabditis elegans*²⁶ have recently been identified. Because, considering the phylogenetic distances, TCP1 is more homologous to these new genes than the cpn60s, we suggest that TCP1 and the new homologues may constitute an independent gene family. Whether TCP1 is classified as a chaperone must await evidence that the heteromeric polypeptide complex purified here can function in protein assembly and/or correct folding. □

Received 10 March; accepted 22 May 1992.

- Silver, L. M., Artzt, K. & Bennett, D. *Cell* **17**, 275–284 (1979).
- Willison, K. et al. *Cell* **57**, 621–632 (1989).
- Willison, K. R., Hynes, G., Davies, P., Goldsborough, A. & Lewis, V. A. *Genet. Res.* **56**, 193–201 (1990).
- Kirchoff, C. & Willison, K. *Nucleic Acids Res.* **18**, 4248 (1990).
- Willison, K. R., Dudley, K. & Potter, J. *Cell* **44**, 727–738 (1986).
- Ahmad, S. & Gupta, R. S. *Biochim. biophys. Acta* **1089**, 253–255 (1990).
- Morita, T., Kubota, H., Gachelin, G., Nozaki, M. & Matsushiro, A. *Biochim. biophys. Acta* **1129**, 96–99 (1991).
- Ursic, D. & Ganetsky, B. *Gene* **68**, 267–274 (1988).
- Ursic, D. & Culbertson, M. R. *Molec. cell. Biol.* **11**, 2629–2640 (1991).
- Gupta, R. S. *Biochem. Int.* **4**, 833–839 (1990).
- Ellis, R. J. *Science* **250**, 954–959 (1990).
- Lingquist, S. & Craig, E. A. *Rev. Genet.* **22**, 631–677 (1988).
- Maekawa, M., O'Brien, D. A., Allen, R. L. & Eddy, E. M. *Biol. Reprod.* **40**, 843–852 (1989).
- Hendrix, R. W. *J. molec. Biol.* **129**, 375–392 (1979).

- Hutchinson, E. G. et al. *EMBO J.* **8**, 1485–1490 (1989).
- Phipps, B. M., Hoffmann, A., Stetter, K. O. & Baumeister, W. *EMBO J.* **10**, 1711–1722 (1991).
- Trent, J. D., Nimmesgern, E., Wall, J. S., Hartl, F.-U. & Horwich, A. L. *Nature* **354**, 490–493 (1991).
- Dahlmann, B. et al. *FEBS Lett.* **251**, 125–131 (1989).
- Harks, S. K., Quinn, A. M. & Hunter, T. *Science* **241**, 42–52 (1988).
- Ellis, R. J. & van der Vies, S. M. A. *Rev. Biochem.* **60**, 321–347 (1991).
- Yaffe, M. B. et al. *Nature* **358**, 245–248 (1992).
- Kim, Y. S., Perdomo, J. & Nordberg, J. *J. biol. Chem.* **266**, 5466–5476 (1991).
- Corbett, J. & Dunn, M. J. *Methods in Molecular Biology: Membrane Methods* (eds Graham, J. & Higgins, J.) (Humana, New Jersey, in the press).
- Mehra, V., Sweetser, D. & Young, R. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7013–7017 (1986).
- Hemmingsen, S. M. et al. *Nature* **333**, 330–334 (1988).
- Waterston, R. et al. *Nature Genet.* **1**, 114–123 (1992).

ACKNOWLEDGEMENTS. We thank M. Dunn for advice with the two-dimensional gels and K. Harrison-Lavoie for discussions. This work was funded by the Cancer Research Campaign.

Cytotoxic T-lymphocyte activation involves a cascade of signalling and adhesion events

Anne M. O'Rourke & Matthew F. Mescher

Division of Membrane Biology, Medical Biology Institute,
11077 N. Torrey Pines Road, La Jolla, California 92037, USA

IN addition to the antigen-specific T-cell receptor (TCR), T cells bear an array of 'accessory' molecules that can contribute to stable adhesion to the antigen-bearing cell¹ and provide costimulatory signals¹⁻⁷. For several of these²⁻⁷, T-cell adhesion to the ligand can be activated by TCR-dependent signalling (a signal from the TCR primes the coreceptor to bind to its ligand). It is unclear whether the individual coreceptors share common mechanisms of priming and cosignalling, and perhaps act in a redundant manner, or whether they act in a distinct way and contribute uniquely to the activation process. We report here the use of isolated alloantigen, class I proteins and fibronectin ligands to show that coreceptors on cytotoxic T lymphocytes are activated sequentially and deliver distinct biochemical signals on binding to their ligands. TCR engagement activates CD8 by a protein tyrosine kinase-dependent pathway, and CD8 then acts as a signal for initiation of polyphosphoinositide hydrolysis on binding to class I. In contrast, activated adhesion to fibronectin does not initiate polyphosphoinositide hydrolysis, but amplifies hydrolysis once it has been initiated. Thus, cytotoxic T-lymphocyte activation involves a TCR-initiated cascade of adhesion and signalling events leading to response.

Fluid-phase anti-TCR monoclonal antibody is not a sufficient stimulus to trigger degranulation by cloned cytotoxic T-lymphocytes (CTL), but primes the cells for CD8-dependent adhesion to class I protein, which results in triggering of degranulation³. To determine which signalling events might be involved in priming CD8 and in mediating its cosignalling for degranulation, phosphatidylinositol (PtdIns) turnover was examined in [³H]myo-inositol-labelled CTL specific for the H-2K^b alloantigen. Target cells⁸ or immobilized K^b stimulated the release of [³H]inositol phosphates (InsP), whereas other (non-antigen) class I proteins did not (Fig. 1a). No InsP release could be detected when cells were treated with fluid-phase anti-TCR antibody alone, under conditions that prime CD8 (ref. 3), but the combination of anti-TCR antibody and class I resulted in InsP production comparable to that found for alloantigen (Fig. 1a).

Stimulated release in response to class I specifically required engagement of the TCR; antibodies against other CTL surface proteins did not initiate release (Fig. 1b). In the presence of anti-TCR monoclonal antibody, activation of the PtdIns pathway occurred with all class I proteins tested, including D^b (Fig. 1), D^d, K^d, K^k and D^k (not shown), and three different anti-CD8 antibodies blocked release of InsP under these conditions (Fig. 1c). These results strongly suggest that engagement of the TCR

FIG. 1 Activated binding of CTL to class I protein initiates [³H]InsP turnover. The indicated quantities of purified class I proteins were immobilized in microtitre wells before blocking with BSA. Turnover of [³H]PtdIns in clone 35 was determined as described below after incubation at 37 °C for 120 min. **a**, c, F23.1 anti-TCR antibody was added in fluid phase at a final concentration of 0.25 µg ml⁻¹ immediately before cells were added to microtitre wells. **b**, All antibodies (0.5 µg ml⁻¹) were added 5 min before addition of cells to wells. Antibodies were: anti-CD8 (53-6.7; 31-68.8), anti-Thy-1 (HO-13.4), anti-H-2K^k (11-4.1), anti-CD45 (I3/2) and anti-Vβ8 TCR (F23.1). **c**, Fluid-phase blocking antibodies were added 15 min before addition of stimulatory F23.1 antibody. Antibodies were: anti-CD8 (a=53-6.7 at 8 µg ml⁻¹, b=31-68.8 at 1:20 dilution of supernatant, c=YTS 169.4 at 1:100 dilution of ascites) or anti-Thy-1 (HO-13.4 and 10 µg ml⁻¹). Treatment of CTL with any of these antibodies in the absence of class I protein did not result in PtdIns turnover, irrespective of the presence of F23.1 antibody. Results are the means and ranges of duplicate samples. Details of the sources of antibodies have been given elsewhere²¹.

METHODS. The cloned CTL used in this study, clones 11, 35 and 30 are all specific for K^b alloantigen, and their derivation and maintenance have been described²². Class I MHC proteins were affinity-purified from tumour cell lysates by monoclonal antibody affinity chromatography²³. Immobilization of class I proteins in microtitre wells, followed by blocking of unbound sites with 1% BSA was as described²³. For determination of [³H]inositol phosphate release, clones were labelled overnight with [³H]myo-inositol in inositol-free supplemented Eagle's basal medium plus 10% (v/v) dialysed FCS, containing 10 µCi ml⁻¹ [³H]myo-inositol (Amersham, S.A. 60-100 Ci mmol⁻¹). [³H]inositol phosphates were extracted at the end of the incubations by removing unbound cells into tubes containing chloroform/methanol (1/2 v/v) plus 0.22 M HCl. Bound cells were extracted by sequential addition of methanol and chloroform/methanol to the wells. This method results in extraction of 100% of cell-associated [³H]phospholipids and inositol phosphates. Combined extracts were then phase-partitioned and the aqueous fraction subjected to Dowex (HCOO⁻) ion-exchange chromatography as described⁸. In all experiments, cells were incubated in RPMI 1640 medium containing 10 mM HEPES and 0.1% (w/v) BSA plus 10 mM LiCl.

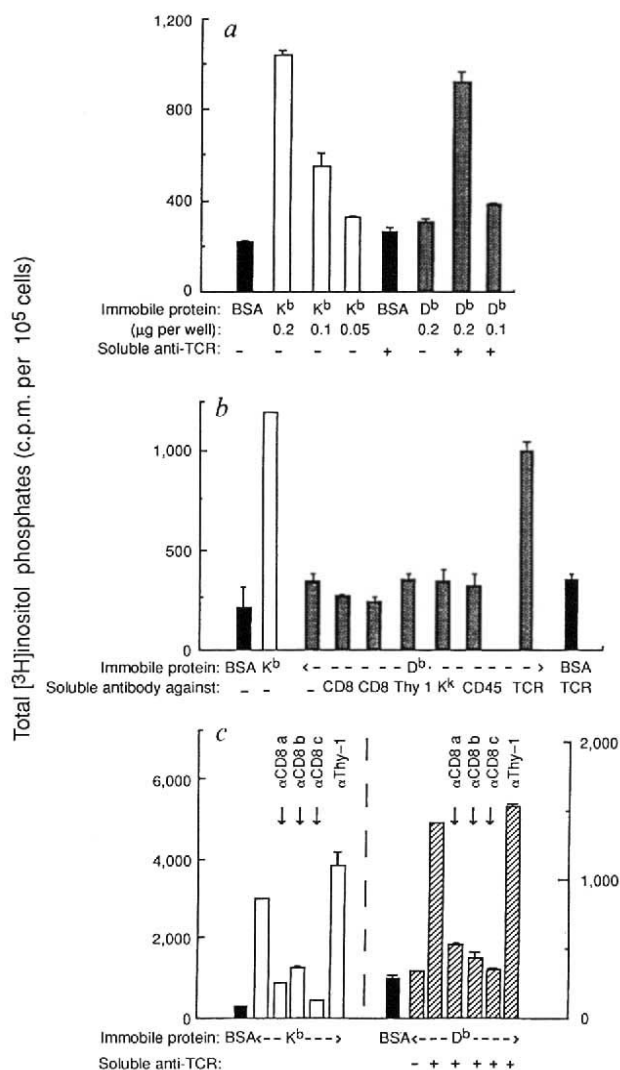
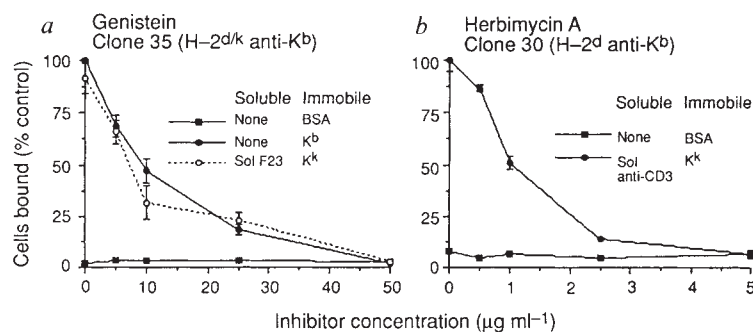


FIG. 2 Tyrosine kinase inhibitors block activated binding of CTL to class I protein. CTL clones were labelled overnight with [methyl- ^3H]thymidine (Dupont NEN; SA 6–10 Ci mmol $^{-1}$) at 2 $\mu\text{Ci ml}^{-1}$, washed and added to wells coated with BSA, K b or K k antigens. CTL (10^5 per well) were treated with the indicated concentrations of genistein (ICN Biochemicals), herbimycin A (Drug Synthesis and Chemistry Branch, National Cancer Institute, NIH) or dimethylsulphoxide vehicle (0.05% final) for 30 min before addition to wells. Where indicated, F23.1 or 2C11-145 (anti-CD3 ϵ) were added at 0.25 $\mu\text{g ml}^{-1}$ or 5 $\mu\text{g ml}^{-1}$, respectively, immediately before initiation of the binding assay. Plates were incubated for 120 min before cells bound were quantified. PBS (100 μl) was added to each well, and the well contents removed by flicking the plate. This was repeated five times before addition of 100 μl per well 3% (w/v) SDS to solubilize cells remaining bound. SDS extracts were then removed and subjected to liquid scintillation counting. Results represent means and the range of duplicate samples.



with fluid-phase antibody does not activate the PtdIns pathway but does prime CD8, and that CD8 binding to class I then activates phospholipase C to initiate PtdIns hydrolysis. That CD8 acts to initiate PtdIns hydrolysis, rather than acting to amplify a low, undetectable level of PtdIns turnover was supported by the absence of detectable InsP production in response to fluid-phase anti-TCR antibody, even when examined over the course of 4 h and in the presence of LiCl to allow accumulation of InsP. Consistent with this, no rise in intracellular free calcium concentration ($[\text{Ca}^{2+}]_i$) can be detected in Fura-2-labelled CTL on addition of fluid-phase anti-TCR antibody (A.M.O. and M.F.M., unpublished results).

Consistent with priming of CD8 occurring before PtdIns pathway activation, CD8-dependent adhesion does not occur on treatment of cells with phorbol myristate acetate (PMA), ionomycin or both, to mimic activation of this pathway (data not shown). In the absence of these signals, TCR-dependent activation of protein tyrosine kinase seemed likely to be the signal for priming of CD8, and the effects of two specific protein tyrosine kinase inhibitors, genistein and herbimycin A, support this. Both compounds effectively block CTL adhesion to class I in the presence of fluid-phase anti-TCR antibody in a dose-dependent manner (Fig. 2). As expected, PtdIns turnover and degranulation were inhibited in parallel, with essentially the same IC_{50} (half-maximal inhibitory concentration) as for inhibition of adhesion (data not shown).

In contrast to class I protein, adhesion of CTL to fibronectin was activated by PMA (Fig. 3). In further contrast, anti-TCR antibody alone activated binding to both ligands, but addition of a second antibody to crosslink further the TCR significantly increased adhesion to fibronectin while inhibiting binding to

class I (Fig. 3a). Effects of the two ligands on PtdIns pathway activation also differed markedly (Fig. 3b). Cells treated with anti-TCR antibody were stimulated to release inositol phosphates when class I was present but not when fibronectin was present, despite the fact that adhesion to fibronectin occurs with this stimulus. Similarly, although PMA activated binding to fibronectin, no InsP release could be detected. Although fluid-phase anti-TCR antibody alone does not stimulate release of InsP, a low level of release is detected on addition of a second, crosslinking antibody. Under these conditions, adhesion to fibronectin stimulated a large increase in the release of [^3H]InsP (Fig. 3b), suggesting that binding to fibronectin cannot initiate PtdIns hydrolysis, but can dramatically amplify InsP release once it has been minimally activated. This was further confirmed by examining InsP release as a function of the amount of anti-TCR antibody added to the cells (Fig. 4a). Fibronectin significantly affected InsP production only under conditions where anti-TCR and second antibody alone could cause a low but detectable level of release (Fig. 4a).

Experiments examining PtdIns hydrolysis in response to co-immobilized alloantigen and fibronectin support the conclusion that TCR and CD8 initiate PtdIns hydrolysis, whereas fibronectin amplifies it. K b alloantigen alone, a ligand for both TCR and CD8, activates PtdIns hydrolysis in a density-dependent manner and this is blocked by anti-CD8 antibodies (Fig. 1). Fibronectin alone does not stimulate release of InsP, but when co-immobilized with alloantigen causes a substantial increase in release (Fig. 4b). Thus, as expected for an amplification mechanism, the effects of fibronectin and alloantigen are highly synergistic and result in levels of release that are not achieved even at high surface densities of alloantigen when it is present

FIG. 3 Activated binding and [^3H]inositol phosphate release by CTL in response to fibronectin or class I protein. a, [Methyl- ^3H]thymidine-labelled clone 35 (10^5 per well) were added to wells blocked with BSA or coated with mouse fibronectin (Telios Pharmaceuticals; 0.5 μg per well) or D b (0.2 μg per well). Cells were treated with F23.1 at 0.5 $\mu\text{g ml}^{-1}$, alone or plus goat anti-mouse immunoglobulin at 5 $\mu\text{g ml}^{-1}$, or 10 nM PMA for 10 min before the start of the assay. Cell binding was determined after 3 h incubation as described for Fig. 2. Examination of the time course for binding has shown that plateau binding level is reached within 1 h and remains constant for at least 4 h. b, [^3H]myo-inositol-labelled clone 35 (10^5 per well) were treated as described for a, and incubated for 120 min before extraction and quantification of [^3H]inositol phosphates as described for Fig. 1.

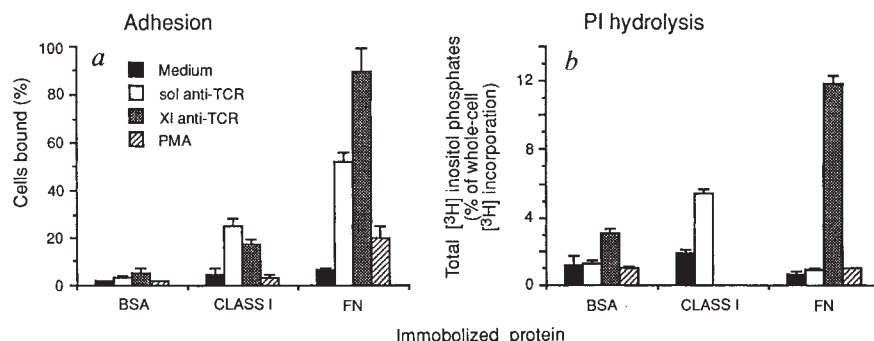
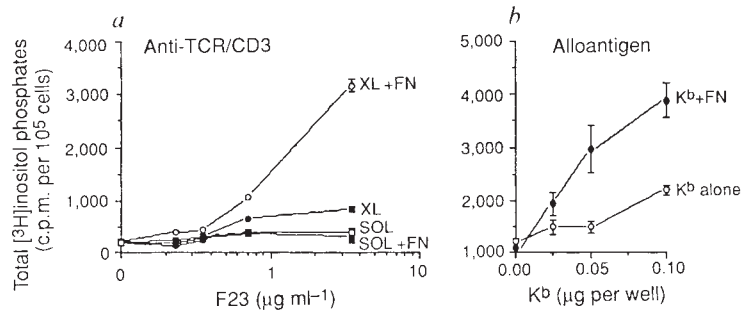


FIG. 4 Activated binding of CTL of fibronectin results in amplification of ongoing PtdIns turnover. *a*, [3 H]inositol-labelled clone 35 (10^5 per well) were treated with the indicated concentration of F23 antibody alone (SOL), or with subsequent addition of $10 \mu\text{g ml}^{-1}$ crosslinking goat anti-mouse immunoglobulin (XL). Cells were then added to wells coated with BSA or with fibronectin at $0.5 \mu\text{g}$ per well and incubated for 120 min before extraction of [3 H]inositol phosphates as described in Fig. 1. *b*, Clone 35 (10^5 per well) were added to wells coated with the indicated concentration of K^b (K^b alone) or to wells coated for 90 min with $0.25 \mu\text{g}$ per well fibronectin, washed with PBS, and then coated with the indicated amount of K^b (K^b +fibronectin). The incubation was continued for 180 min before [3 H]inositol phosphates were extracted. Incubations for shorter or longer periods did not result in detectable PtdIns turnover in response to soluble anti-TCR antibody plus fibronectin. Bars indicate the range of duplicate samples.



alone (Fig. 4*b* and data not shown).

Activation of a protein tyrosine kinase (PTK) is the earliest event detected after TCR engagement^{9,10}, and seems likely to provide the signal that leads to priming to CD8 (Fig. 2). This must involve some additional step(s), because neither the α - nor β -chains of murine CD8 have tyrosines in the cytoplasmic regions^{11,12}. Adhesion to alloantigen also depends on PTK activity (Fig. 2*a*), lending further support to the suggestion that adhesion to antigen is an actively signalled process involving TCR-dependent priming of CD8 for binding. The p56^{lck} PTK associated with CD8 (ref. 13) is the likely candidate for provision of a signal for activation of phospholipase leading to PtdIns hydrolysis. CD8- p56^{lck} might directly phosphorylate phospholipase C to activate it^{14,15}. Alternatively, feedback to the TCR-CD3 complex may occur, possibly involving phosphorylation of the ζ chain of the TCR complex¹⁶, creating a binding site for the SH2 *src*-homologous domains of phospholipase C; a possibility supported by the recent demonstration that phospholipase $\text{C}\gamma$ -1 can be co-immunoprecipitated with the TCR complex¹⁷.

The results suggest that CTL adhesion to fibronectin is not fully activated until after PtdIns hydrolysis has been initiated. After this, adhesion to fibronectin acts to amplify further this pathway. Consistent with the results shown here for InsP release, co-immobilized fibronectin substantially augments the level of CTL degranulation in response to alloantigen (B. Ybarrodo *et al.*, manuscript in preparation). Both very late antigen integrins VLA-4 and VLA-5 can bind fibronectin⁴ and preliminary results indicate that the effects described here result predominantly from binding by VLA-5. The mechanism by which VLA mediates PtdIns amplification is not known, but recent reports suggest that PTKs may also be involved in signalling by integrins^{18,19}.

Thus, antigen, class I and fibronectin ligands act together to activate CTL through a cascade of distinct adhesion and signalling events initiated by the TCR. In addition, preliminary experiments using murine intercellular adhesion molecule²⁰ indicate that the requirements for activation of leukocyte function-associated antigen (LFA-1)-dependent adhesion to this ligand, and the resulting cosignalling effects, differ from those for either class I or fibronectin (A.M.O. *et al.*, unpublished results). Thus, the multiple 'accessory' interactions involving adhesion/coreceptor function seem not to be redundant systems. Instead, they act sequentially to strengthen adhesion to the antigen-bearing surface, and deliver distinct transmembrane signals on binding to their ligands. The existence of multiple, receptor-activated adhesion/coreceptor systems that can act in a cascade to initiate responses brings an additional level of complexity to the understanding of functional activation resulting from cell-cell interactions. □

- O'Rourke, A. M., Rogers, J. & Mescher, M. F. *Nature* **346**, 187-189 (1990).
- Shimizu, Y., van Seventer, G. A., Horgan, K. J. & Shaw, S. *Nature* **345**, 250-253 (1990).
- van Seventer, G. A., Shimizu, Y., Horgan, K. J. & Shaw, S. *J. Immunol.* **144**, 4579-4586 (1990).
- Shimizu, Y., van Seventer, G. A., Horgan, K. J. & Shaw, S. *J. Immunol.* **145**, 59-67 (1990).
- van Seventer, G. A. *et al. J. exp. Med.* **174**, 901-913 (1991).
- O'Rourke, A. M. & Mescher, M. F. *J. Biol. Chem.* **263**, 18594-18597 (1988).
- June, C. H., Fletcher, M. C., Ledbetter, J. A. & Samelson, L. E. *J. Immunol.* **144**, 1591-1599 (1990).
- Mustelin, T., Coggeshall, K. M., Isakov, N. & Altman, A. *Science* **247**, 1584-1587 (1990).
- Nakauchi, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 5126-5130 (1985).
- Panaccio, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 6874-6878 (1987).
- Veilleille, A., Zuniga-Pflucker, J. C., Bolen, J. B. & Kruisbeek, A. M. *J. exp. Med.* **170**, 1671-1680 (1989).
- Secrist, J. P., Karnitz, L. & Abraham, R. T. *J. Biol. Chem.* **266**, 12135-12139 (1991).
- Weiss, A., Koretzky, G., Schatzman, R. C. & Kadlecsek, T. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5484-5488 (1991).
- Barber, E. K., Dasgupta, J. D., Schlossman, S. F., Trevillyan, J. M. & Rudd, C. E. *Proc. natn. Acad. Sci. U.S.A.* **86**, 3277-3281 (1989).
- Dasgupta, J. D. *et al. J. exp. Med.* **175**, 285-288 (1992).
- Kornberg, L. J., Earp, H. S., Turner, C. E., Prockop, C. & Juliano, R. L. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8392-8396 (1991).
- Nojima, Y., Rothstein, D. M., Sugita, K., Schlossman, S. F. & Morimoto, C. *J. exp. Med.* **175**, 1045-1053 (1992).
- Siu, G., Hedrick, S. M. & Brian, A. A. *J. Immunol.* **143**, 3813-3820 (1989).
- Kane, K. P. & Mescher, M. F. *J. Immunol.* **144**, 824-829 (1990).
- Kane, K. P., Sherman, L. A. & Mescher, M. F. *J. Immunol.* **142**, 4153-4160 (1989).
- Kane, K. P., Champoux, P. & Mescher, M. F. *Molec. Immunol.* **26**, 759-768 (1989).

ACKNOWLEDGEMENTS. We thank J. Rogers and P. Champoux for technical assistance. This work was supported by grants from the NIH (M.F.M.), and a Special fellowship from the Leukemia Society of America (A.M.O.R.).

Superantigen implicated in dependence of HIV-1 replication in T cells on TCR $\text{V}\beta$ expression

Jeffrey Laurence, Andrew S. Hoddtsen & David N. Posnett*

Department of Medicine, Cornell University Medical College, 1300 York Avenue, Box 56, New York, New York 10021, USA

IN the pathogenesis of AIDS it is not yet understood whether the small fraction of CD4^+ T cells ($\sim 1\%$) infected with the human immunodeficiency virus (HIV) are randomly targeted or not. Here we present evidence that human CD4^+ T-cell lines expressing selected T-cell antigen receptor $\text{V}\beta$ gene products can all be infected *in vitro* with HIV-1, but give markedly different titres of HIV-1 virion production. For example, $\text{V}\beta 12$ T-cell lines from several unrelated donors reproducibly yielded up to 100-fold more *gag* gene product (p24^{gag} antigen) than $\text{V}\beta 6.7\alpha$ lines. This is consistent with a superantigen effect, because the $\text{V}\beta$ selectivity was observed with several divergent HIV-1 isolates, was dependent on antigen-presenting cells and on major histocompatibility complex (MHC) class II but was not MHC class II-restricted. The *in vivo* significance of these findings is supported by the preferential stimulation of $\text{V}\beta 12^+$ T cells by freshly obtained irradiated antigen-presenting cells from some HIV-1-seropositive but not

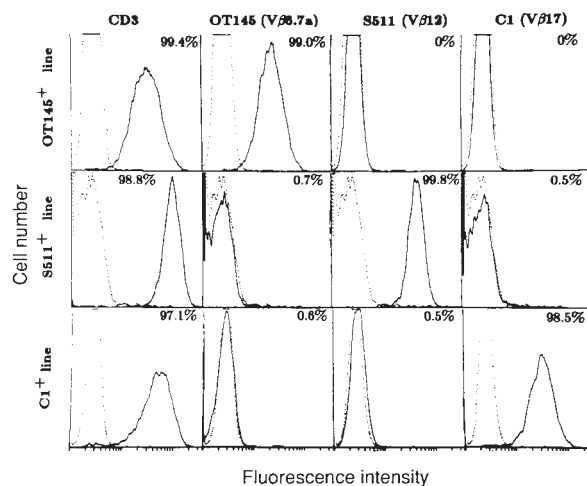
Received 18 February; accepted 28 May 1992.

- Springer, T. A. *Nature* **346**, 425-434 (1990).
- Dustin, M. L. & Springer, T. A. *Nature* **341**, 619-624 (1989).

* To whom correspondence should be addressed.

FIG. 1 Phenotype of 3-week-old polyclonal cell lines expressing selected $V\beta$ gene products. Three cell lines derived from donor S.F. are stained by indirect immunofluorescence with the monoclonal antibody indicated. The percentage of positive cells are shown in each panel. The dotted line represents a negative control antibody.

METHODS. Cell lines were derived in parallel from fresh PBL of normal donors by incubating with monoclonal antibodies raised against TCR V gene products for 1 h, washing the cells and separating them with goat anti-mouse immunoglobulin-coated magnetic dynabeads (Dyna). Cell lines were maintained with IL-2 twice a week and with irradiated allogeneic periodate-treated non-T PBLs (APC) every two weeks. Monoclonal antibodies are S511- $V\beta 12$ (refs 23, 24), OT145- $V\beta 6.7a$ (ref. 25), C1- $V\beta 17$ (ref. 26) and Ti3a- $V\beta 8$ (ref. 27).



HIV-1-negative donors. Moreover, cells from patients positive for viral antigen (gp120) were enriched in the $V\beta 12$ subpopulation. $V\beta 12^+$ T cells were not deleted in AIDS patients, however, raising the possibility that a variety of mechanisms contribute to T-cell depletion. Our results indicate that a superantigen targets a subpopulation of $CD4^+$ cells for viral replication.

Superantigens are typically recognized by T cells expressing

particular T-cell antigen receptor (TCR) $V\beta$ genes¹. To test the possibility that HIV-1, like murine mammary tumour viruses¹⁻⁵, is associated with superantigen activity, we used a new method for rapid preparation of polyclonal human T-cell lines expressing selected $V\beta$. T cells that were >99% positive for $V\beta 12$, $V\beta 6.7a$, $V\beta 17$ or $V\beta 8$ expression (verified using monoclonal antibodies S511, OT145, C1 and Ti3a, respectively) were pre-

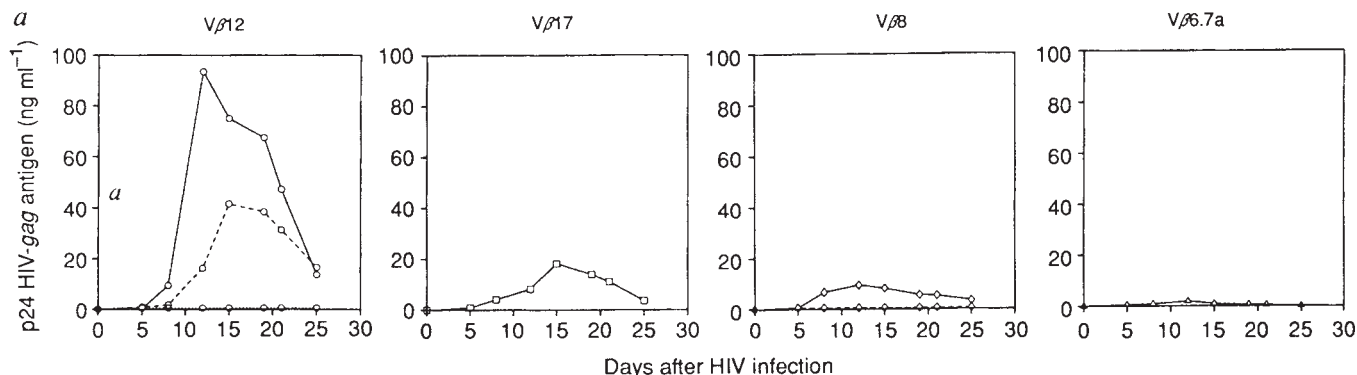
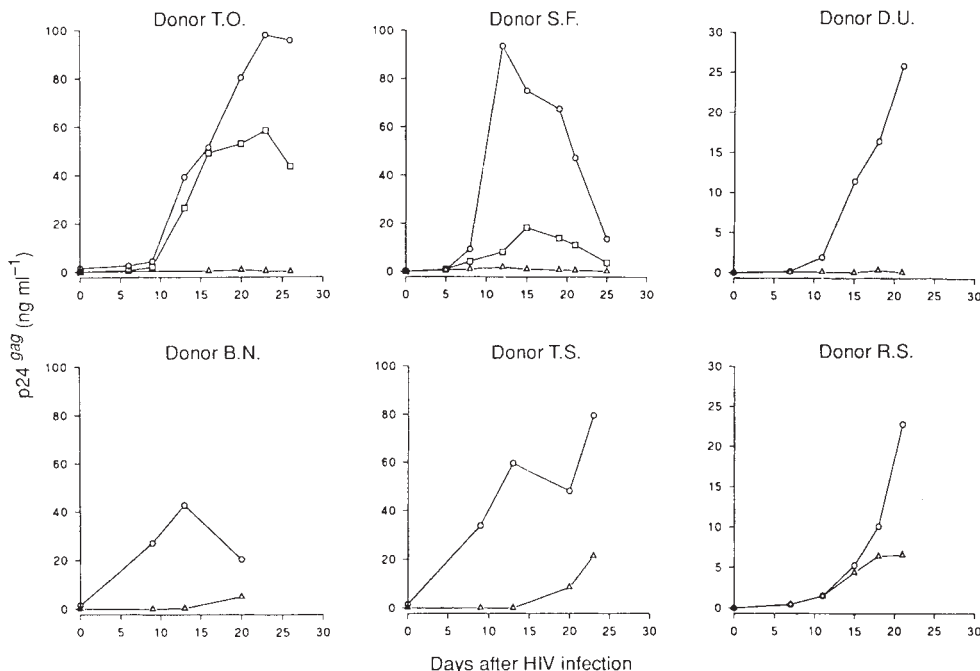


FIG. 2 HIV-1 replication in parallel cultures expressing different TCR $V\beta$ products. **a**, parallel cell lines were obtained from donor S.F. Viral doses were 1,000 TCID-50 (tissue culture infectious doses, 50%) (unbroken line), 100 TCID-50 (dashed line), and 10 TCID-50 (dotted line). **b**, Differential replication of HIV-1 in $V\beta$ -selected cell lines from six unrelated HLA-mismatched normal donors. Cell lines expressed $V\beta 12$ (circles), $V\beta 17$ (squares), $V\beta 8$ (diamonds), and $V\beta 6.7a$ (triangles).

METHODS. 5×10^5 cells per ml were plated in 2-ml wells on day 0 and exposed to TIIB isolate for 2 h at 37 °C. Cells were then washed three times and replated in culture medium with 10% IL-2. Irradiated periodate-treated APCs were added on day 1. Supernatant was harvested for p24 ELISA assay²⁸ every 3-4 days, and half of the culture volume was replaced with fresh medium and IL-2.



pared from several donors (Fig. 1). Cells were exposed to varying titres of the TIIB isolate of HIV-1 for two hours at 37 °C, washed and maintained with antigen-presenting cells (APCs) and interleukin-2 (IL-2). All T-cell lines had measurable amounts of p24^{gag} antigen after day 5 of culture and were therefore infected with HIV-1. This was confirmed by polymerase chain reaction (PCR) with gag-specific primers. But the degree of HIV-1 replication varied greatly (Fig. 2), with up to 100-fold differences observed between Vβ12- and Vβ6.7a-expressing cell lines. Vβ17 and Vβ8 cell lines supported intermediate levels of HIV-1 replication. T-cell activation is not required for HIV-1 infection⁶ but activation of T cells promotes a high level of HIV replication. If a superantigen were involved in activating Vβ12 cells, but not Vβ6.7a cells, similar results would be expected in cell lines derived from unrelated HLA-mismatched donors because superantigen recognition is typically not MHC-restricted¹. Vβ12 cell lines from six different donors supported the highest levels of virion production, Vβ17 intermediate levels, and Vβ6.7a the lowest levels (Fig. 2b). The cell lines were simultaneously derived from each donor and contained similar ratios of CD4/CD8 cells (2:1), and activation antigens such as IL-2Rα (CD25), 4F2 (ref. 7) and HLA-DR were comparably expressed. In several instances, CD4 and CD8 cells were separated. Only CD4 S511⁺ cells, and not the CD4 OT145⁺ or the CD8 sublines, gave a high level of HIV-1 replication (Fig. 3a). Thus, depletion of CD8 cells from the Vβ6.7a cells did not restore high-level HIV-1 replication, mitigating the possibility that CD8⁺ T cells capable of suppressing HIV-1 replication⁸ were more active in OT145⁺ than in S511⁺ cultures.

Cultures to which irradiated APCs had not been added supported only minimal HIV-1 replication (Fig. 3a). Moreover, if MHC class II monoclonal antibodies were added to the cultures, they blocked enhanced HIV-1 replication in Vβ12 cells. In one experiment we measured p24^{gag} in 12-day HIV-1 infected cultures as 1,692 pg ml⁻¹ (Vβ12 cells) and 291 pg ml⁻¹ (Vβ6.7a cells). A control antibody, OKM1, had no effect, but MHC class II antibody decreased p24^{gag} expression to 358 pg ml⁻¹ in Vβ12 cells (80% suppression), consistent with presentation of a putative superantigen by MHC class II.

HIV-1 may initially infect MHC class II-positive T cells or irradiated APC in these cultures, producing a superantigen capable of activating selected T-cell subpopulations (including S511⁺ T cells), which would then readily integrate extrachromosomal proviral DNA and efficiently direct production of infec-

tious virions^{9,10}. To ensure that Vβ6.7a cells had the capacity to replicate HIV-1 to high levels they were maximally activated by adding phorbol myristate acetate (PMA) and phytohaemagglutinin (PHA) (Fig. 3b). Vβ6.7a cells could now support high levels of HIV-1 replication identical to those in Vβ12 cells. Vβ12 cells were not pre-activated independently of HIV-1 exposure, because pre-exposure (day 0) cultures had identical activation antigen phenotypes and identical baseline DNA synthesis rates, as determined by ³H-thymidine uptake. In contrast, 12 days after HIV-1 exposure, Vβ12 cells expressed 6-fold more CD25, 5-fold more HLA-DR, and 2–3-fold less CD4, CD3 and TCR than day-12 Vβ6.7a cells. These results suggest activation of Vβ12 but not Vβ6.7a cells by an HIV-1-associated superantigen, and are consistent with previous results on HIV-induced down-regulation of CD4 and CD3/TCR expression¹¹.

To determine whether HIV-1 isolates other than TIIB were also associated with the observed superantigen effect, we used the divergent BaL strain¹². Here also Vβ12 cells supported much higher HIV-1 replication than Vβ6.7a cells (Fig. 3c). In addition, a new patient isolate (Mar) gave HIV-1 replication more than two logs higher in Vβ12 cells than in Vβ6.7a cells.

In mice, *in vivo* administration of a superantigen induces varying degrees of clonal deletion and clonal anergy^{13–18}. Superantigen exposure can also lead to anergy of antigen-specific human T-cell clones *in vitro*¹⁹. As Vβ12 CD4 cells appear to promote HIV-1 replication *in vitro*, this population might be expected to be selectively eliminated *in vivo*⁴. Alternatively, they might represent a small subset of infected T cells that remain a reservoir for latent virus¹⁰. In 10 HIV-seropositive patients with varying degrees of CD4 T-cell depletion, normal percentages of Vβ12⁺ CD4⁺ cells were observed compared with normal controls (2.0 ± 0.7 versus 2.1 ± 1.0), and Vβ12⁺ CD4⁺ cells were as abundant as Vβ12⁺ CD8⁺ cells (Fig. 4a, b). Vβ17 and Vβ6.7a positive cells were also normally represented. Two-colour analysis was also done with anti-gp120 monoclonal antibody as a marker for HIV-1 infection, and anti-Vβ antibody (Fig. 4c). A striking enrichment of gp120⁺ cells was observed in the Vβ12 subpopulation, but not among Vβ5.1, Vβ6.7a, Vα2.3 or Vβ8 cells in 2 out of 4 HIV-1 seropositive patients. This finding has been confirmed by PCR data which demonstrate that levels of HIV-1 gag DNA in Vβ12 cells are higher than in Vβ6.7a cells (D.N.P. *et al.*, manuscript in preparation). To test whether an HIV-1 superantigen with specificity for Vβ12 gene products is produced *in vivo*, irradiated T-depleted PBL from HIV-1⁺

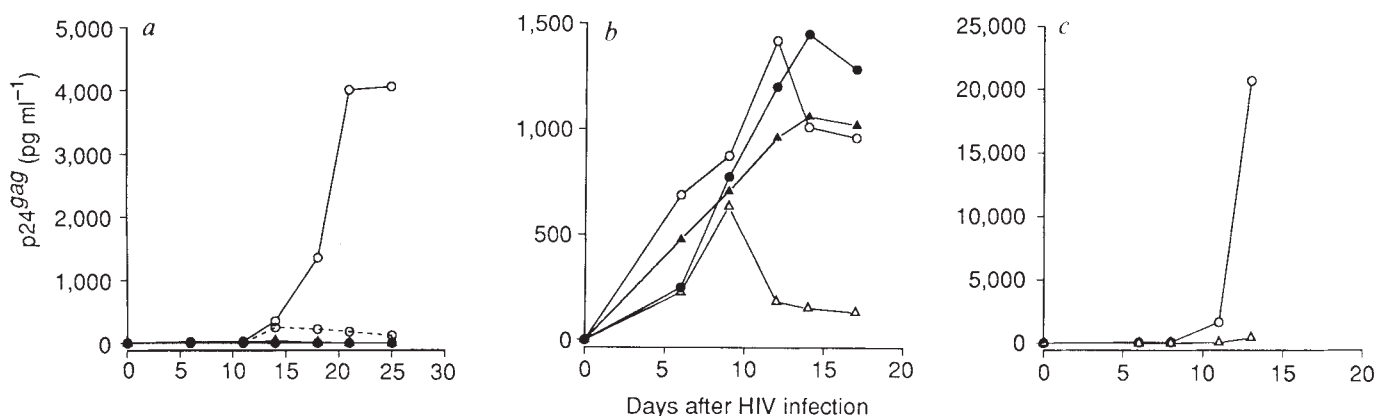


FIG. 3 HIV-1 infection of CD4⁺ and CD8⁺ sublines, the role of APCs, the role of T-cell activation and the effect of different HIV-1 isolates were tested. In a, CD4 and CD8 sublines of the Vβ12 and Vβ6.7a cell lines from donor T0 were obtained by negative selection with Dynabeads. Only the CD4 Vβ12 subline (circles) and not the CD8 Vβ12 cell line (filled circles), the CD4 Vβ6.7a cell line (triangles), or the CD8 Vβ6.7a cell line (filled triangles) supported high-level HIV-1 replication (34 pg μl⁻¹ of p24 with CD4 Vβ6.7a cells). Omission of APC on day 1 of culture resulted in much-

decreased HIV-1 replication (dashed line). In b, CD4⁺ Vβ12 (open and filled circles) and CD4⁺ Vβ6.7a (open and filled triangles) cell lines were either treated with PHA (5 μg ml⁻¹) and PMA (10 ng ml⁻¹) (solid symbols) or untreated (open symbols), demonstrating that activated CD4⁺ Vβ6.7a cells were equally effective at sustaining HIV-1 replication as CD4⁺ Vβ12 cells. In c, the HIV-1 isolate BaL was used, demonstrating that CD4 Vβ12 cells (circles) support high level HIV-1 replication, in contrast to CD4 Vβ6.7a cells (triangles).

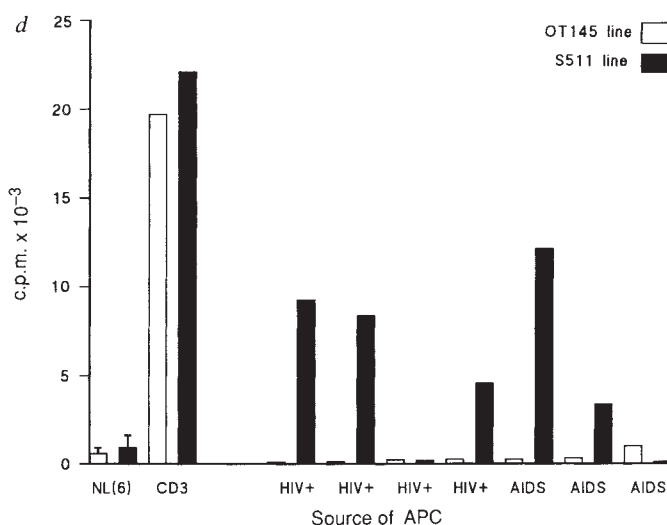
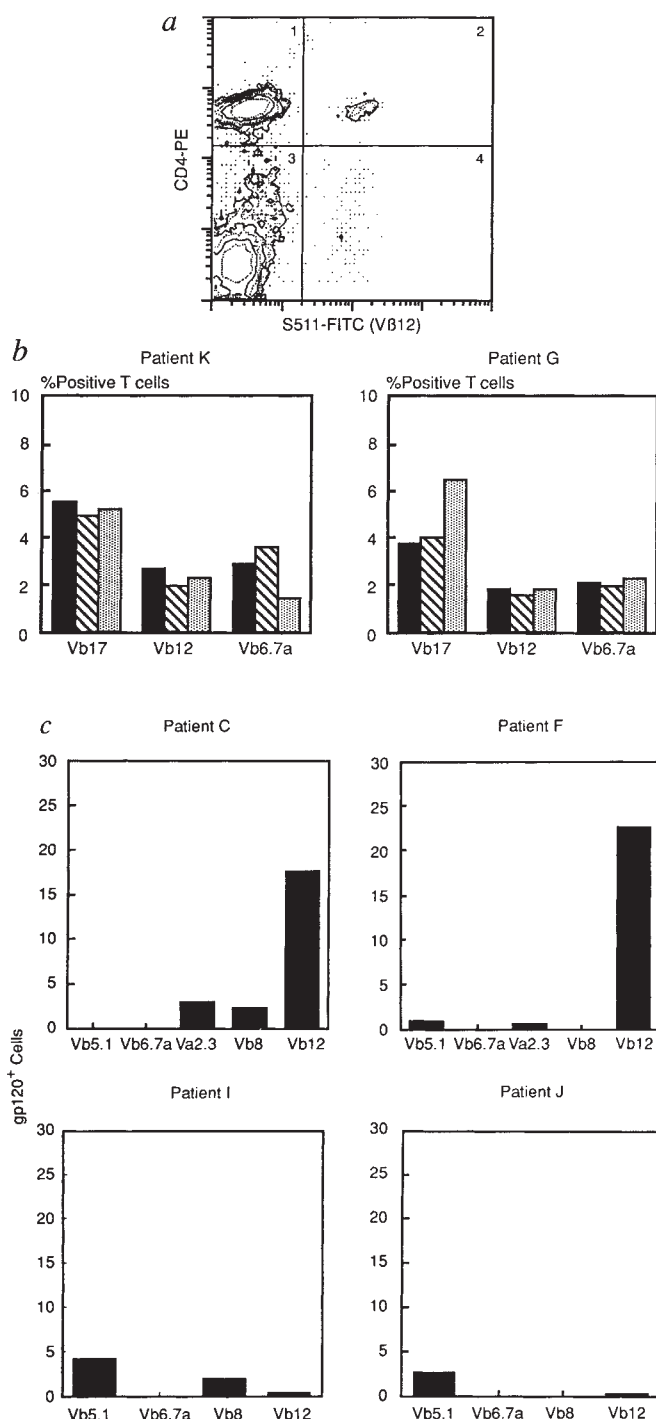


FIG. 4 *In vivo* V β expression and expression of a superantigen in HIV⁺ patients. *a*, Peripheral blood T cells from an HIV-1⁺ patient are stained with the S511-fluorescein isothiocyanate (FITC) and CD4-phycoerythrin (PE) to demonstrate a discrete population of double-positive cells in quadrant 2. *b*, Two-colour fluorescence data on two representative patients are expressed as percentages of V β 17, V β 12 or V β 6.7a positive cells of CD3 (black bars), of CD4 (striped bars), or of CD8 cells (stippled bars). Patient G had symptomatic AIDS (<10 CD4 cells per mm³) and patient K was asymptomatic (<200 CD4 cells per mm³). Similar data were obtained from 8 other patients with CD4 blood counts ranging from 700 to <10 cells per mm³. *c*, Peripheral blood T cells were analysed by 2-colour FACS with anti-gp120 antibody versus directly FITC-conjugated V β -specific antibody. Percentages of gp120⁺ cells in the various V β subpopulations are shown. *d*, T-cell-depleted peripheral blood mononuclear cells from 7 patients (HIV⁺ or AIDS) and 6 controls (NL) were irradiated with 5,000 rads and co-cultured with in-parallel-derived V β 12 (solid bars) or V β 6.7a (open bars) cell lines at a 1:1 cell ratio. ³H-thymidine incorporation was measured after 3 days. Alloresponses to non-T cells from normal donors were not observed with the culture conditions used.

METHODS. For immunofluorescence, T cells were first stained with an anti-V β antibody, then goat anti-mouse F(ab')₂-FITC (Tago Inc.), then quenched with excess negative control IgG1 mouse antibody, and stained with PE-conjugated CD3, CD4 or CD8 antibodies (Olympus). For staining with anti-gp120 antibody (Dupont), the staining order was anti-gp120, followed by goat anti-mouse-PE, quenching with mouse IgG1 and FITC-labelled anti-V β antibodies (T Cell Sciences, Inc.).

patients were used as stimulators for established V β cell lines. As shown in Fig. 4d, 5 out of 7 patient samples induced V β 12 but not V β 6.7a cell lines to proliferate, whereas cells from six control donors did not activate either under the conditions used, and monoclonal antibody against CD3 activated both. This is consistent with expression of a V β 12-specific superantigen *in vivo* on APC contained within the non-T cell fraction.

In summary, V β 12 cells are preferential targets for HIV-1 replication. Our results support earlier speculation that virus-encoded superantigens activate target T cells to facilitate viral replication⁵. The concept of a preferential target for HIV-1 replication is not unique to retrovirus biology. A single TCR α/β specificity appears to mark the primary T cell for murine leukaemia virus RadLV induction of cell proliferation and enhanced virion production²⁰ and exogenous murine mammary

tumour virus requires a superantigen-targeted V β 14 subset to infect its host³.

In contrast with a recent report of *in vivo* V β usage measured by PCR (ref. 21), V β 17 was not decreased in our patients (Fig. 4b), regardless of the CD4 T-cell count or their clinical status. A possible pitfall with seriously ill patients is that the observed T-cell repertoire skewing may not be due to HIV, but to other opportunistic pathogens. The combined *in vitro* and *in vivo* studies presented here suggest that V β selectivity is probably due to HIV-1. It is still unclear what role a viral superantigen may play in rendering targeted T cells anergic and/or susceptible to deletion by apoptosis²². The absence of preferential V β 12 or V β 17 T-cell depletion raises the possibility that mechanisms other than superantigen-induced deletion contribute to CD4 T-cell depletion. However, if HIV-1 produces a superantigen

that contributes to the mechanism of infection of CD4 T cells, novel therapeutic approaches can be designed to inhibit the activity of the superantigen. □

Received 16 April; accepted 4 June 1992.

- Herman, A., Kappler, J. W., Marrack, P. & Pullen, A. M. *Rev. Immun.* **9**, 745–772 (1991).
- Marrack, P., Kushnir, E. & Kappler, J. *Nature* **349**, 524–526 (1991).
- Golovkina, T. V., Chervinsky, A., Dudley, J. P. & Ross, S. R. *Cell* **69**, 637–645 (1992).
- Janeway, C. A. *Nature* **349**, 459–460 (1991).
- Coffin, J. M. *Science* **255**, 411–413 (1992).
- Zack, J. A. *et al. Cell* **61**, 213–222 (1990).
- Haynes, B. F. *et al. J. Immun.* **126**, 1409–1414 (1981).
- Walker, C. M., Moody, D. J., Staties, D. P. & Levy, J. A. *Science* **234**, 1563–1566 (1986).
- Stevenson, M., Stanwick, T. L., Dempsey, M. P. & Lamonica, C. A. *EMBO J.* **9**, 1551–1560 (1990).
- Bukrinsky, M. I., Stanwick, T. L., Dempsey, M. P. & Stevenson, M. *Science* **254**, 423–427 (1991).
- Laurence, J., Friedman, S. M., Chartash, E. K., Crow, M. K. & Posnett, D. N. *J. clin. Invest.* **83**, 1843–1848 (1989).
- Gartner, S. *et al. Science* **233**, 215–219 (1986).

- Webb, S., Morris, C. & Sprent, J. *Cell* **63**, 1249–1256 (1990).
- Rammensee, H. G., Kroschewski, E. & Frangoulis, B. *Nature* **339**, 541–544 (1989).
- Kawabe, Y. & Ochi, A. *J. exp. Med.* **172**, 1065–1070 (1990).
- Blackman, M. *et al. Nature* **345**, 540–542 (1990).
- Rellahan, B. L., Jones, L. A., Kruisbeek, A. M., Fry, A. M. & Matis, L. S. *J. exp. Med.* **172**, 1091–1100 (1990).
- Ignatowicz, L., Kappler, J. & Marrack, P. *J. exp. Med.* **175**, 917–923 (1992).
- O'Hehir, R. E. & Lamb, J. R. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8884–8888 (1990).
- O'Neill, H. *Cell. Immun.* **136**, 54–61 (1991).
- Imberti, L., Sottini, A., Bettinardi, A., Puoti, M. & Primi, D. *Science* **254**, 860–862 (1991).
- Groux, H. *et al. J. exp. Med.* **175**, 331–340 (1992).
- Bigler, R. D., Fischer, D., Wang, C. Y., Rinnooy-Kan, E. & Kunkel, H. G. *J. exp. Med.* **158**, 1000–1005 (1983).
- Kappler, J. *et al. Science* **244**, 811–813 (1989).
- Prasher, Y., Li, Y., Kubinec, J. S., Jones, N. & Posnett, D. N. *J. Immun.* **147**, 3441–3444 (1991).
- Friedman, S. M. *et al. J. exp. Med.* **174**, 891–900 (1991).
- Acuto, O. *et al. J. exp. Med.* **161**, 1326–1343 (1985).
- Laurence, J., Sikder, S., Kulkosky, J., Miller, P. & Ts'o, P. O. P. *J. Virol.* **65**, 213–219 (1991).

ACKNOWLEDGEMENTS. This work was supported in part by grants from the National Institutes of Health and the US Army Medical Research Acquisitions Activity. A.S.H. is a fellow of the Cancer Research Institute/Samuel Freeman Charitable Trust.

Retinoblastoma protein switches the E2F site from positive to negative element

Steven J. Weintraub, Cheryl A. Prater & Douglas C. Dean*

Departments of Medicine and Cell Biology and Physiology, Washington University School of Medicine, 660 S. Euclid, St Louis, Missouri 63110, USA

ORIGINALLY E2F sites were identified as elements in the promoters of adenovirus early genes that are necessary for activation of these genes by the early protein E1a (ref. 1). E2F promoter elements have been shown to be important for transcriptional activation of several genes critical for progression through the cell cycle^{2–4}. During the G1 phase of the cell cycle, the E2F protein forms a complex with the cell-cycle protein Rb (ref. 5) and it has been suggested that this binding of Rb to E2F inactivates E2F (ref. 5). Here we show that Rb–E2F is an active complex that, when bound to the E2F site, inhibits the activity of other promoter elements and thus silences transcription. We propose that the ability of this complex to inhibit transcription is integral to the function of Rb and provide evidence that E2F is a positive element in the absence of an active form of Rb. It has been shown that binding of Rb to E2F depends on the phosphorylation state of Rb (only the underphosphorylated form binds)⁵ and that the phosphorylation state of Rb changes during progression through the cell cycle^{6,7}. We therefore suggest that the E2F site alternates between a positive and negative element with the phosphorylation/dephosphorylation cycle of Rb. This cyclic activity may be responsible for activating and then inhibiting genes during the cell cycle.

We first investigated the contribution of E2F to the activity of the promoter for the adenovirus early gene *E1a*. In addition to E2F sites, this promoter contains several strong enhancers⁸, but its basal activity is relatively low (data not shown), suggesting that the promoter may contain a negative element. The role of the E2F protein in *E1a* promoter activity was examined in transfection assays in which a competitor plasmid containing E2F binding sites was cotransfected with the plasmid pE1aCAT, which contains the *E1a* promoter fused to the gene for chloramphenicol acetyltransferase (CAT). This competitor should bind and sequester E2F, thus preventing it from interacting with the *E1a* promoter. The results show that pE1aCAT expression is activated in the presence of this competitor (Fig. 1), indicating that the E2F sites act to inhibit *E1a* promoter activity. As a control, the activity of the Rous sarcoma virus long terminal

repeat (pRSVCAT) was unaffected by the competitor (Fig. 1).

To determine whether E2F sites could act as negative elements in the context of other promoters, we cloned E2F sites into a simple promoter containing only a TATA box and an ATF transcription factor site (Fig. 2a). These E2F sites inhibited the activity of this construct (Fig. 2b and c). To confirm that inhibition by E2F is a general phenomenon, the effect of cloning E2F sites into the simian virus-40 early gene promoter/enhancer was examined. As in the simple promoter described above, E2F sites inhibited the activity of this relatively complex promoter (Fig. 2b and c).

The adenovirus protein E1a can activate E2F sites⁹ and this is thought to occur because E1a dissociates complexes formed between E2F and other cellular proteins. This releases what may be an active form of E2F (ref. 10). In agreement with previous studies, we show that E2F sites activate transcription in the

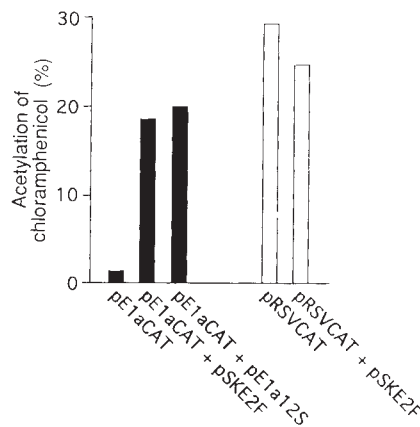
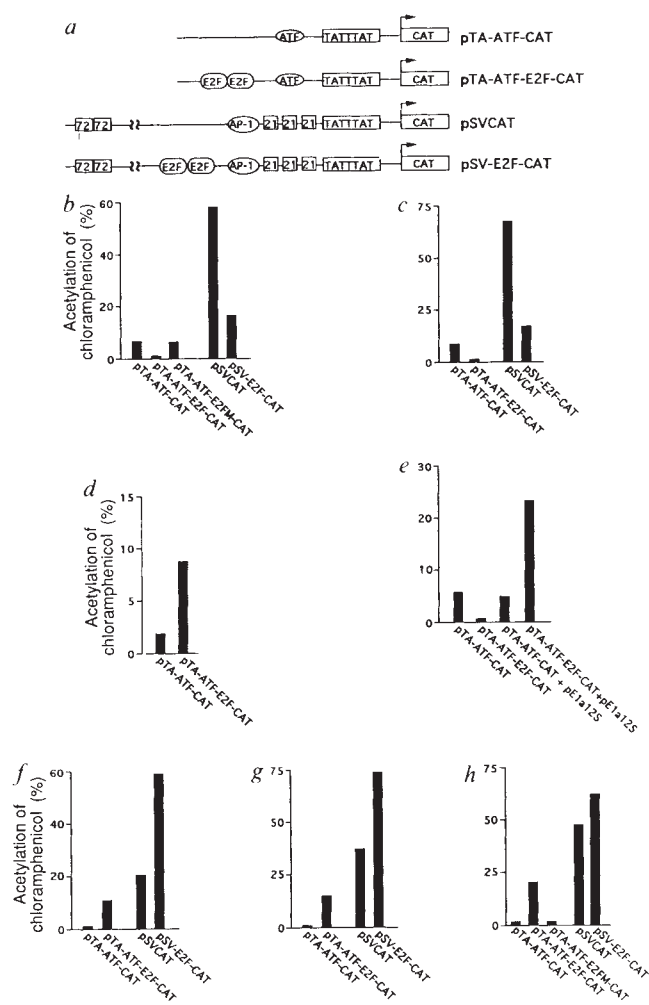


FIG. 1 The E2F site can inhibit transcription. A competitor plasmid containing E2F sites (pSKE2F) was used to sequester cellular E2F and prevent it from interacting with the *E1a* promoter (pE1aCAT) in transfection assays.

METHODS. pE1aCAT contains the *E1a* promoter from position +61 to –498 bp fused to the CAT gene¹³. pRSVCAT contains the Rous sarcoma virus long terminal repeat fused to the CAT gene. To construct pSKE2F, oligonucleotides containing the E2F binding sites from the adenovirus E2a promoter¹⁴ (5'-AGCTTGTTCGCGCTTAATTTGAGAAAGGCGCGAACTA-GTCA-3') were synthesized such that after annealing a *Hind*III site is present on each end. Only the coding strand is shown and E2F sites are underlined. The annealed oligonucleotides were cloned into Bluescript SK that had been digested with *Hind*III. Transfections were done in the mink lung epithelial cell line CCL64 using the calcium phosphate method¹⁵ and contained 5 µg reporter plasmid along with 75 µg pSKE2F or the parent vector Bluescript SK. Where indicated, 2 µg pE1a12S, which encodes the 243-amino-acid form of E1a¹³, was cotransfected. Thirty-six hours after transfection, cells were collected and CAT activity was determined using 100 µg protein in 3-h reactions¹⁵. Per cent acetylation of chloramphenicol was determined by thin layer chromatography followed by scintillation counting. Results are an average of duplicate samples and are representative of at least 3 separate experiments.

* To whom correspondence should be addressed.



presence of E1a: in the 293 cell line, which stably expresses E1a, E2F sites were positive elements (Fig. 2*d*), and they acted as positive elements when an expression vector encoding the 243-amino-acid form of E1a was cotransfected into mouse L cells (Fig. 2*e*).

Taken together, these results indicate not only that free E2F is a transcriptional activator (as already suggested¹⁰), but that a complex between E2F and a cellular protein(s) (which is dissociated by E1a) acts to inhibit transcription.

One of the proteins that binds to E2F is the retinoblastoma

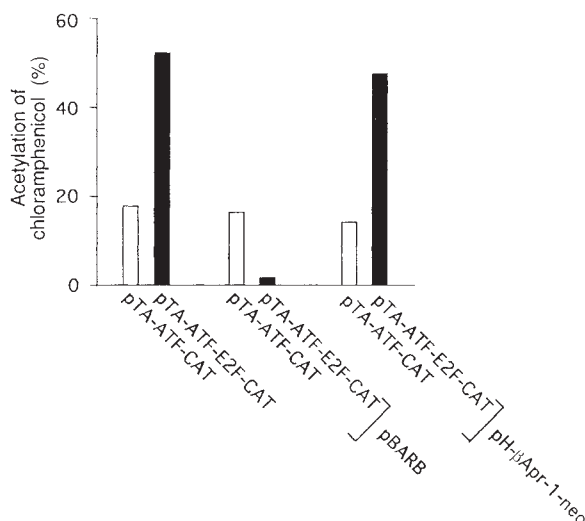


FIG. 2 E2F sites act as negative elements in Rb(+) cell lines and positive elements in the presence of E1a and in Rb(-) cell lines. The effect of E2F sites on the activity of two heterologous promoters was compared in a series of transfections into Rb(+) (mouse L cells and the mink lung epithelial cell line CCL64) and Rb(-) cell lines (the osteosarcoma cell line SAOS-2 (ref. 16), the bladder carcinoma cell line HTB-9 (ref. 17), and the cervical carcinoma cell line C33A¹⁸). The 293 cell line stably expresses E1a. *a*, Diagram of reporter plasmids; *b*, mink lung epithelial cells; *c*, L cells; *d*, 293 cells; *e*, L cells; *f*, SAOS-2 cells; *g*, HTB-9; *h*, C33A cells.

METHODS. Construction of pTA-CAT and pTA-ATF-CAT has been described¹³. pSVCAT (Promega Biotech.) contains the SV40 early gene promoter (the TATA element, the 21-bp repeats, and the AP-1 site) and enhancers (72-bp repeats). E2F sites from pSKE2F were removed with *HindIII* digestion and cloned into the *HindIII* site of pTA-ATF-CAT to create pTA-ATF-E2F-CAT. The E2F sites were blunt-ended with T4 polymerase and cloned into the *BglII* site in pSVCAT, which had also been blunt-ended, to create pSV-E2F-CAT. Mutant E2F site oligonucleotides were cloned into pTA-ATF-CAT to create pTA-ATF-E2FM-CAT as in Fig. 1:



E2F sites are underlined and mutations are shown above the sequence. Transfections were as described for Fig. 1.

protein Rb (refs 5, 11). Overexpression of Rb in cells has been shown to inhibit promoter activity¹², but the specific element that mediates this effect was not identified nor was it determined whether Rb functions as a true negative regulator or simply to downregulate a positive factor(s). Nevertheless, these results indicate that Rb can have a negative effect on transcription and they are consistent with the possibility that Rb-E2F could be a complex that inhibits transcription.

Constructs described in Fig. 2a were transfected into cell lines expressing only a mutant form of Rb (SAOS-2, HTB-9 and

FIG. 3. Transfection of an Rb expression vector into Rb(-) cell lines causes the E2F site to switch from positive to negative element. pBARB contains the human β -actin promoter fused to the human Rb cDNA and pH-BAP-1-neo is the parent vector that lacks Rb cDNA¹⁷. Similar results were observed with a vector where the SV40 early promoter/enhancer was fused to the Rb cDNA (J3 Ω HRbC and the parent vector J3 Ω)¹⁹.

METHODS. Equal molar concentrations of pBARB (3 μ g) and pH- β APr-1-neo (2 μ g) were transfected into C33A cells along with the indicated reporter plasmids as described in Fig. 1 legend.

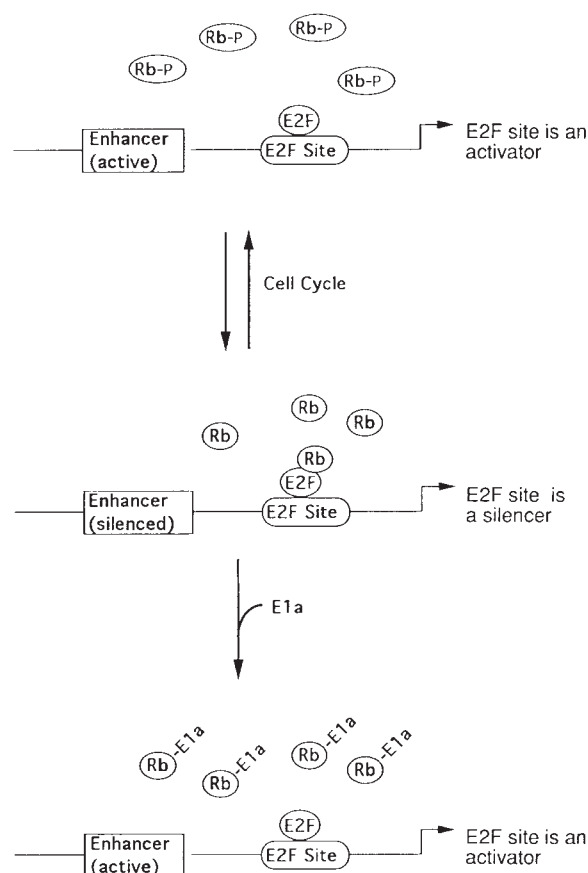


FIG. 4 Proposed role of Rb in E2F site activity. E2F sites act as negative elements that inhibit the activity of other promoter elements in the presence of Rb but are positive elements in the absence of an active form of Rb (Figs 2 and 3). E2F-site activity oscillates between a positive and negative element according to the phosphorylation state of Rb during the cell cycle. E1a prevents binding of Rb to E2F (ref. 10) and E2F sites switch from negative to positive elements in the presence of E1a. Thus E1a disrupts the regulatory effect of the phosphorylation/dephosphorylation cycle of Rb. Rb-P represents the phosphorylated form of Rb, and E1a-Rb is a complex of Rb with E1a²⁰.

C33A) to determine whether the inhibitory activity of the E2F site is dependent upon Rb. In contrast to Rb(+) cell lines in which E2F sites were negative elements (Fig. 2b and c), E2F sites stimulated transcription in Rb(-) cell lines (Fig. 2f-h), suggesting that this inhibition depends on Rb. Mutation of the E2F sites in pTA-ATF-E2F-CAT (pTA-ATF-E2FM-CAT) prevents binding of nuclear proteins in gel retardation assays (data not shown), and eliminates the E2F site-dependent inhibition in Rb(+) cells (Fig. 2b) and the E2F-site-dependent activation in Rb(-) cells (Fig. 2h), demonstrating that functional E2F sites are necessary for the results shown in Fig. 2.

To determine whether the difference in E2F site activity in Rb(+) and Rb(-) cells is due solely to Rb, we cotransfected an Rb expression vector along with reporter constructs into an Rb(-) cell line. In the presence of the Rb expression vector, the E2F sites were inhibitory (the same activity observed in Rb(+) cell lines; Fig. 2b and c) (Fig. 3). As an additional control, the Rb expression vector had no effect on pRSVCAT (data not shown). These results confirm that E2F sites are negative elements in the presence of Rb and positive elements in its absence. Not only is this evidence that Rb-E2F is an active complex that inhibits transcription, it also shows directly that E2F is a positive

element in the absence of Rb.

It has been proposed that binding of Rb to E2F simply inactivates E2F (ref. 5), but our suggestion that Rb-E2F is an active complex that inhibits transcription is probably critical to the function of Rb. It would allow Rb to have a profound effect on promoter activity by switching E2F sites from positive to negative elements. It is doubtful that simple inactivation of E2F could effectively inhibit the activity of promoters containing E2F sites because several of these contain strong enhancers.

The binding of Rb to E2F is controlled by phosphorylation of Rb (ref. 5). The phosphorylation state of Rb varies with the cell cycle: it is relatively underphosphorylated early in G1 but is hyperphosphorylated as the cell approaches S phase and remains phosphorylated until late in mitosis, when it is dephosphorylated^{6,7}. As only the underphosphorylated form of Rb binds to E2F (ref. 5), our results are consistent with a model in which E2F sites in the promoters of genes important for cell proliferation are responsible for sequentially activating and inhibiting expression of these genes during the cell cycle (Fig. 4). The E1a oncoprotein is able to disrupt this regulatory effect of the phosphorylation/dephosphorylation cycle of Rb by preventing the binding of Rb to E2F (Fig. 4). □

Received 1 April; accepted 14 May 1992.

- Kovesdi, I., Reichel, R. & Nevins, J. R. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2180-2184 (1987).
- Blake, M. C. & Azizkhan, J. C. *Molec. cell. Biol.* **9**, 4994-5002 (1989).
- Thalmeier, K., Synovzik, H., Mertz, R., Winnacker, E. L. & Lipp, M. *Genes Dev.* **3**, 527-536 (1989).
- Mudryj, M., Hiebert, S. W. & Nevins, J. R. *EMBO J.* **9**, 2179-2184 (1990).
- Chellappan, S. P., Hiebert, S., Mudryj, M., Horowitz, J. M. & Nevins, J. R. *Cell* **65**, 1053-1061 (1991).
- Buchkovich, K., Duffy, L. A. & Harlow, E. *Cell* **58**, 1097-1105 (1989).
- Ludlow, J. W., Shon, J., Pipas, J. M., Livingston, D. M. & DeCaprio, J. A. *Cell* **60**, 387-396 (1990).
- Hearing, P. & Shenk, T. *Cell* **33**, 695-703 (1983).
- Hiebert, S. W., Blake, M., Azizkhan, J. & Nevins, J. R. *J. Virol.* **65**, 3547-3552 (1991).
- Bagchi, S., Raychaudhuri, P. & Nevins, J. R. *Cell* **62**, 659-669 (1990).
- Chittenden, T., Livingston, D. M. & Kaelin, W. G. *Cell* **65**, 1073-1082 (1991).
- Robbins, P. D., Horowitz, J. M. & Mulligan, R. C. *Nature* **346**, 668-671 (1990).

- Weintraub, S. J. & Dean, D. C. *Molec. cell. Biol.* **12**, 512-517 (1992).
- Yee, A. S., Reichel, R., Kovesdi, I. & Nevins, J. R. *EMBO J.* **6**, 2061-2068 (1987).
- Rosen, G. D., Birkenmeier, T. M. & Dean, D. C. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4094-4098 (1991).
- Shew, J. Y. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6-10 (1990).
- Takahashi, R. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5257-5261 (1991).
- Scheffner, M., Munger, K., Byrne, J. C. & Howley, P. M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5523-5527 (1991).
- Bernards, R. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6474-6478 (1989).
- Whyte, P. et al. *Nature* **334**, 124-129 (1988).

ACKNOWLEDGEMENTS. We thank R. Takahashi for pBARB and PH-βApr-1-neo and R. Weinberg for J3Ω and J3ΩHRbc. This work was supported by grants from the NIH. S.J.W. is an Allen and Hansburys fellow.

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without interruptions of the text, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, who are chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are typeset from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose, and to mention availability of these data.

Once a manuscript is accepted for publication, contributors will receive galley proofs in about 4 weeks. *Nature's* staff will edit manuscripts with a view to brevity and clarity, so contributors should check their proofs carefully. Manuscripts are generally published 2–3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review articles survey recent developments in a field. Most are commissioned but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50–80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few subheadings of two or three words.

Letters are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned but suggestions can be made to the commentary editor in the form of a one-page synopsis. Commentaries are normally between one and four pages of *Nature*.

News and Views articles inform non-specialist readers about new scientific advances, sometimes in the form of a conference report.

Most are commissioned but proposals can be made in advance to the News and Views editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words and 5 references.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and three copies are required, each accompanied by artwork. If photographs are included, four sets of originals are required; for line drawings, one set of originals and three good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Relevant manuscripts in press or submitted for publication elsewhere should be included with each copy of a submitted manuscript, and clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their manuscript numbers.

Titles say what the paper is about with the minimum of technical terminology and in fewer than 80 characters in the case of Articles and Letters. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs. Numbering of sequences should be in the left-hand margin only, with a single space between rows.

Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome.

Colour artwork. A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent the publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference, including unrefereed abstracts, should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963–65). The first and last page numbers are included; reference to books should include publisher, place and date.

Abbreviations, symbols, units and greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used except for changed addresses.

Acknowledgements are brief; grant and contribution numbers are not allowed.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1137 National Press Building, Washington, DC 20045, USA. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax (VAT). □